

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
2 May 2002 (02.05.2002)

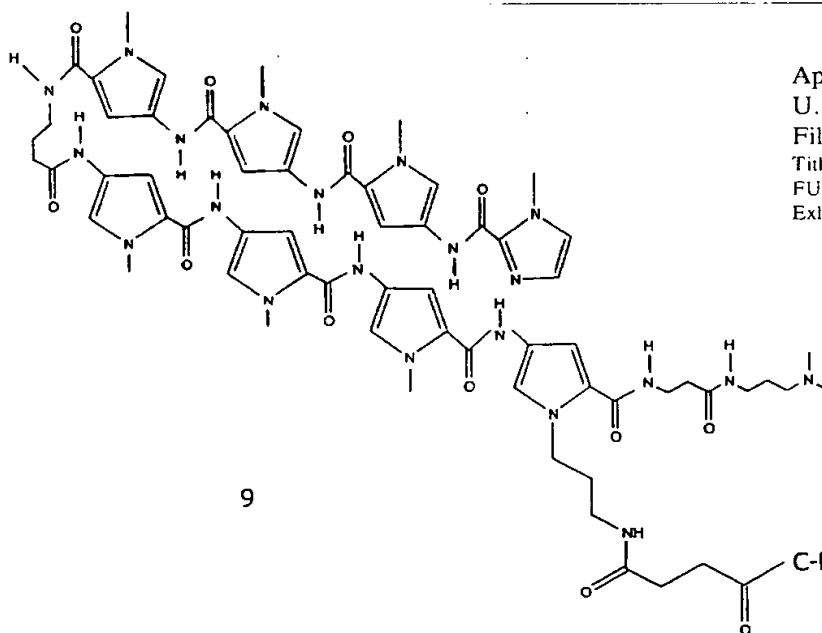
PCT

(10) International Publication Number
WO 02/34295 A1

- (51) International Patent Classification⁷: **A61K 47/48**,
49/00
- (21) International Application Number: **PCT/US00/29617**
- (22) International Filing Date: 27 October 2000 (27.10.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (71) Applicants (for all designated States except US): **MEMORIAL SLOAN-KETTERING CANCER CENTER** [US/US], 400 East 64th Street, New York, NY 10021 (US); **CALIFORNIA INSTITUTE OF TECHNOLOGY** [US/US], 1200 East California Boulevard, Pasadena, CA 91125 (US)
- (72) Inventors: and
- (75) Inventors/Applicants (for US only): **DERVAN, Peter** [US/US], 1235 Saint Albans Road, San Marino, CA 91108 (US); **MAPP, Anna** [US/US], 1700 Shadford Road, Ann Arbor, MI 48104 (US); **PTASINE, Mark** [US/US], 9 East 79th Street, New York, NY 10021 (US); **ANSARI,**
- (74) Agent: **LEHNHARDT, Susan, K.**; Frommer Lawrence & Haug LLP, 745 Fifth Avenue, New York, NY 10151 (US).
- (81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published: — with international search report

[Continued on next page]

(54) Title: SYNTHETIC REGULATORY COMPOUNDS



Applicant: Ulrich Laemmli
U.S. Serial No.: 09/892,085
Filed: June 26, 2001
Title: MODULATION OF CHROMOSOME
FUNCTION BY CHROMATIN REMODELING...
Exhibit 2

(57) Abstract: This invention provides novel synthetic regulatory compounds that comprise a nucleic acid binding moiety, a linker, and a regulatory moiety, compositions comprising such compounds, methods of designing and synthesizing such compounds, methods of screening such compounds to identify those having the desired regulatory activity, and methods of using such compounds to prevent or treat disease in plants and animals, including humans. These compounds, and compositions containing them, have multiple applications, including use in human and animal medicine and in agriculture.



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TITLE OF THE INVENTION

SYNTHETIC REGULATORY COMPOUNDS

CROSS-REFERENCE TO RELATED APPLICATIONS

- 5 This application claims the benefit of and priority to United States provisional patent application serial number 60/161,545, filed October 26, 1999.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

The United States government may have certain rights to this invention pursuant to National Institute of Health grant no. GM-27681.

10 FIELD OF THE INVENTION

This invention relates to synthetic regulatory compounds comprising a double-stranded nucleic acid binding moiety, a linker, and a regulatory moiety, the synthesis and testing of such compounds, and applications therefor.

BACKGROUND OF THE INVENTION

- 15 The following description of the background of the invention is provided to aid in understanding the invention, but is not admitted to be or to describe prior art to the invention.

- The regulation of gene expression is critical to the growth, development, proliferation, and maintenance of all living cells and organisms. In most cases, the positive or negative regulation of genes is under the control of signal transduction cascades that transmit information from the cell surface to the nucleus. Signal-transduction cascades are generally triggered by ligands which may be small molecules, soluble peptides, extracellular matrix, adhesive proteins attached to cell surfaces of neighboring or migrating cells, and even metabolic intermediates. In most cases, ligands interact with a membrane bound, or sometimes soluble intracellular, receptor, thus triggering a cascade of events that ultimately either stimulate or inhibit the activity of the mRNA-synthesizing machinery at one or more genes. Such reprogramming of gene-expression leads to an appropriate cellular response to the stimuli. Based on current understanding, almost all such signals converge and mediate their function through transcriptional activators and/or repressors, although some environmental stimuli, such as nutrient deprivation, directly affect the stability of certain components of the transcriptional machinery.
- 20
- 25
- 30

Because of the importance of gene transcription or gene expression to living cells, manipulating the process is of extreme interest. Compounds that fundamentally alter the activity of the transcriptional machinery itself, for example, by inhibiting the elongation process, would be potent transcriptional modulators, and it is gene-specific regulation that is the goal of many development programs. One approach to gene-specific transcriptional regulation has been to develop molecules that block activator-DNA or repressor-DNA interactions and thereby regulate transcription artificially. Several approaches in this vein are being investigated. One such approach involves protein nucleic acid (PNAs) which are oligomers that contain the standard purine and pyrimidine bases of an oligonucleotide but contain a simple amide-based backbone as opposed to the sugar-phosphate backbone found in nucleic acids. Nielsen (1997) Chem. Eur. J. 3:505-508. PNAs have been shown to bind with very high affinity to single-stranded DNAs and RNAs. It has also been shown that a PNA complementary to one strand of a DNA duplex will invade the double helix, pair with its complementary strand. Footer et al. (1996) Biochem. 35:10673-10679. PNA binding can abolish PNA protein interactions in the same region, and additionally PNAs have been employed as antisense agents.

Another class of such molecules is oligonucleotides that are capable of promoting "triple helix" formation. Yet another class of molecules are the so-called "polyamides" developed by Dervan and co-workers. See, e.g., Dervan et al. (1999) Curr. Opin. Chem. Biol. 3:688-693. Polyamides having high sequence specificity and association constants have been developed based on a "code" by which a given pair of substituted or unsubstituted imidazole /pyrrole pair can be selected to bind to particular nucleotide base pairs in the minor groove of double stranded DNA. Sequence specific polyamides that adhere with proteins that bind via the major-groove of double stranded DNA have also been developed. Bremer et al. (1998) Chem. Biol. 5:119-133.

In addition to developing molecules that interfere with the association of activators and repressors with the cognate target sequences, another approach involves small molecules that modulate (positively or negatively) interactions between proteins involved in the regulation of transcription. A second approach to regulating transcription of a desired gene may involve mediating protein-protein

interaction. To date, efforts in this area have involved cell-based genetic approaches.

For example, the so-called "two-hybrid assay" (Fields et al. (1989) Nature 340:245-246) is based on the observation that in many promoter contexts, the DNA
5 binding and activation domains of an activator protein function more or less independently of one another Brent, R. & Ptashne, M. (1985) Cell 43:729-736; Keegan, L. et al (1986) Science 231:691-704.

Nonetheless, these domains require functional association in proximity of the promoter. For instance, if the activation and DNA binding domains of the yeast
10 Gal4 protein are severed and expressed in a yeast strain deleted for wild-type GAL4, no transcription of genes under the control of the GAL4 promoter occurs. However, in genes encoding two other proteins that interact with one another are fused to the DNAs encoding the severed GAL4 domains, activator activity is reconstituted and the target gene can be transcribed. Ma, J. & Ptashne, M. (1988) Cell 55:443-446.
15 Other similar systems, each of which requires the intracellular expression of chimeric gene constructs, are known. Vidal et al. (1996) Proc. Natl. Acad. Sci. USA 93:10321-10326; Leanna et al. (1996) Nucl. Acids Res. 24:3341-3347; Huang et al. (1997) Proc. Natl. Acad. Sci. USA 94:13396-13401; and Hu et al. (1990) Science 250:1400-1403.

20 Despite these approaches, however, at present there exists no class of synthetic, cell-permeable compounds that can regulate the expression of a specific gene.

BRIEF SUMMARY OF THE INVENTION

It is the object of this invention to provide a novel class of synthetic
25 regulatory compounds, which are preferably cell permeable, as well as compositions comprising such compounds, methods of synthesizing such compounds, methods of screening to identify such compounds, and methods of using such compounds.

Thus, in one aspect, the invention concerns synthetic regulatory compounds, each of which comprises at least one nucleic acid binding moiety, at least one
30 regulatory moiety, and at least one linker connecting the nucleic acid binding moiety(ies) to the regulatory moiety(ies). By "synthetic" is meant any compound wherein the at least one of the particular nucleic acid binding moiety(ies) and at least

one of the particular regulatory moiety(ies) are not found in the same molecule in nature, i.e., in a wild-type animal or plant. By "non-natural" is meant that the compound (e.g., a peptide) is not naturally occurring.

A "nucleic acid binding moiety" refers to any compound that binds to a nucleic acid molecule, be it single- or double-stranded DNA or RNA (with double-stranded DNA being preferred), under desired conditions. What constitutes "desired conditions" will vary depending upon application, but in general refers to reaction conditions such as temperature, pH, solvent, ionic strength, the presence or absence of chaotropic agents, reactant concentrations, etc. to be encountered in the eventual intended application of the compound. In the context of synthetic regulatory compounds to be used as drugs, for example, preferred desired conditions are physiological conditions and, again, these will vary depending upon the particular animal, plant, and environment being considered, but in any event may be determined by one ordinarily skilled in the art. The particular conditions used (for example, the reactions conditions used to conduct in vitro screening or other testing) need not be equivalent to those actually found in a cell for example; however, such conditions should sufficiently represent or approximate those likely to be encountered in the eventual application (e.g., the conditions in a human cancer cell, when the compound is a pharmaceutical intended to treat or prevent such cancer) such that meaningful data can be generated.

Preferably, a nucleic acid binding moiety of a compound according to the invention will specifically interact with a target nucleotide sequence. A "target nucleotide sequence" refers to a specific sequence of nucleotides, and is typically represented in the 5' to 3' direction using standard single letter notation, where "A" represents adenine, "G" represents guanine, "T" represents thymine, and "C" represents cytosine, and "U" represents uracil. As used herein, a "target nucleotide sequence" within a double-stranded nucleic acid molecule preferably comprises a sequence greater than 3 but preferably less than about 20 nucleotides in length.

The target sequence, in general, is defined by the nucleotide sequence on one of the strands of the double-stranded nucleic acid. Such sequences are pre-selected or pre-determined, and are thus referred to as targets. While such sequences comprise a specific sequence of nucleotides, it will be appreciated that such

sequence can include a different nucleotide at the same position, i.e., is degenerate at that position, with respect to one or more positions in the particular sequence.

Degenerate bases can be represented by any suitable nomenclature, for example, that which is described in World Intellectual Property Organization Standard ST.25

5 (1998), Appendix 2. Typically, when a nucleic acid binding moiety of a synthetic regulatory compound of the invention specifically binds to its target nucleotide sequence, it does so in a manner that does not compete for binding site interaction with endogenous compounds. This can be accomplished in any suitable manner, for example, by selecting a target nucleotide sequence that is adjacent or proximal to
10 (e.g., within about 500 bases, preferably about 200 bases, and even more preferably within about 100 or fewer bases of) the binding site for a DNA-binding protein of the transcriptional machinery.

By "regulatory element" is meant any cis element of defined nucleotide sequence that can be identified in a nucleic acid molecule and which associates with
15 a DNA-binding protein of the transcriptional machinery or a protein involved in providing chromatin structure. Such elements include promoters and enhancers. A "promoter" is the minimum sequence necessary to initiate transcription of a target gene. An "enhancer" is a cis-acting sequence that increases the utilization of a eukaryotic promoter. Preferred cis elements to be targeted by the nucleic acid
20 binding moieties of the invention's synthetic regulatory compounds are those that occur endogenously in association with the gene whose transcription is to be regulated.

Other embodiments include chimeric reporter constructs that comprise a promoter or other regulatory element not naturally associated with a particular gene.
25 Transcription from any promoter can be regulated by a synthetic regulatory compound. As such, promoters from which transcription can be initiated by any RNA polymerase can be targeted. Suitable RNA polymerases include, but are not limited to, RNA polymerase I (which transcribes ribosomal RNA (rRNA) in eukaryotic cells), RNA polymerase II (which transcribes messenger RNA (mRNA)
30 in eukaryotic cells), and RNA polymerase III (which transcribes transfer RNA (tRNA) in eukaryotic cells).

The interaction between the nucleic acid binding moiety can occur at different regions in the nucleic acid molecule. For example, when dsDNA in the B form contains the target nucleotide sequence, the nucleic acid binding moiety preferably interacts with the DNA via minor and/or major groove interactions.

5 In preferred embodiments, the nucleic acid binding moiety is a molecular scaffold that spatially arrays a plurality of hydrogen bond donors and acceptors in a manner that allows the formation of hydrogen bonds with the corresponding hydrogen bond acceptors and donors in the target nucleotide sequence when the nucleic acid binding moiety interacts with a nucleic acid molecule (e.g., dsDNA) containing the target nucleotide sequence. In addition, or alternatively, the molecular scaffold can provide moieties that allow specific electrostatic and/or van der Waals interactions between units of the scaffold and atoms in the target nucleic acid molecule. Preferred nucleic acid binding moieties that can act as such molecular scaffolds include protein nucleic acids (PNAs) and oligonucleotides.

15 Another preferred class of nucleic acid binding moiety useful in the practice of this invention can be represented as follows: $-Q_1-Z_1-Q_2-Z_2-\dots-Q_m-Z_m-$, wherein each of Q_1, Q_2, \dots, Q_m is independently selected from a heteroaromatic moiety and $(CH_2)_p$, wherein p is an integer between 1 and 10, inclusive, and is preferably 1, 2, or 3; wherein each of Z_1, Z_2, \dots, Z_m is independently selected from the group consisting of a covalent bond and a linking group; and m is between 1 and 20, inclusive, with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or 16 being preferred. In certain preferred embodiments, at least one of Q_1, Q_2, \dots, Q_m is a heteroaromatic moiety, for example, a substituted or unsubstituted imidazole or pyrrole moiety. In particularly preferred embodiments, at least about 50%, 60%, 70%, 80%, or more of Q_1, Q_2, \dots, Q_m are a heteroaromatic moieties, particularly substituted or unsubstituted imidazole or pyrrole moieties. In additional or alternative preferred embodiments, at least one of Z_1, Z_2, \dots, Z_m is a linking group having between 1 and 10, inclusive, and preferably 2, 3, 4, or 5, backbone atoms. In particularly preferred embodiments, each of Z_1, Z_2, \dots, Z_m is a carboxamide group.

25 Such nucleic acid binding moieties, wherein the vast majority of the hydrogen bond donors and acceptors are contained within moieties (e.g., heteroaromatic moieties

such as substituted or unsubstituted imidazoles or pyrroles) linked by carboxamide bonds, are referred to as polyamide nucleic acid binding moieties.

Polyamides can be designed to assume one of several alternative conformations upon base-specific interaction with a nucleic acid molecule, including hairpin, H-pin, slipped, overlapped, and cyclic conformations. Such conformations include intermolecular 2:1 binding motifs between dsDNA molecules comprising a corresponding target sequence, as well as intramolecular 2:1 binding motifs between dsDNA molecules comprising a corresponding target sequence.

Preferably, a nucleic acid binding moiety included in a synthetic regulatory compound of the invention has, under desired conditions, a binding specificity for its corresponding target nucleotide sequence of at least about two, and preferably at least about 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100 or more, as compared to a mismatch target sequence. A "mismatch" target sequence refers to a target nucleotide sequence in which one nucleotide is different at a particular position, as compared to the target nucleotide sequence.

Preferably, a nucleic acid binding moiety included in a synthetic regulatory compound of the invention also has, under desired conditions, at least about submicromolar, and preferably at least about nanomolar or picomolar binding affinity for its target nucleotide sequence. Binding affinity can be determined by any suitable technique including, but not limited to, quantitative DNase I footprint analysis. Association constants (K_a) for preferred nucleic acid binding moieties are at least about 10^6 M^{-1} , 10^7 M^{-1} , 10^8 M^{-1} , 10^9 M^{-1} , 10^{10} M^{-1} , 10^{11} M^{-1} , or 10^{12} M^{-1} .

Other embodiments of this aspect concern synthetic regulatory compounds that comprise two or more nucleic acid binding moieties associated in a manner that allows each to retain its nucleic acid binding function, for example, through the use of a flexible linker. One representative class of synthetic regulatory compounds comprising a plurality (e.g., 2, 3, 4, or more) of nucleic acid binding moieties include those wherein two such moieties are tethered to one another via a linker, and the linker is attached to another linker (or, alternatively, is a dendrimeric linker having three or more functional sites) that is attached to a regulatory moiety. Another such class concerns compounds wherein the nucleic acid binding moieties are each separately attached via a linker to the regulatory moiety. Yet another class

concerns compounds wherein the nucleic acid binding moieties are tethered via a linker and one of them is also linked via a separate linker to a regulatory moiety. Many other configurations of this sort are encompassed herein. These include, but are not limited to, compounds containing multiple regulatory moieties (e.g., 2, 3, 4, 5 5, 6, 7, 8, 9, 10, or more), and compounds containing multiple ~~regulatory moieties~~ and a plurality (e.g., 2, 3, 4, or more) of nucleic acid binding moieties.

Synthetic regulatory compounds that comprise a plurality of nucleic acid binding moieties typically will have greater target sequence specificity than a synthetic regulatory compound containing fewer nucleic acid binding moieties. In 10 addition, the compounds bind to both in the minor and major grooves of dsDNA, and provide groups that can be modified by other molecules, for example, by linkers to which one or more regulatory moiety can be attached. Similarly, synthetic regulatory compounds that comprise a plurality of regulatory moieties can interact with more than one component of the transcription or chromatin structure 15 machinery.

In order to regulate, or modulate, expression of a target gene, a synthetic regulatory compound according to the invention also contains, in addition to a nucleic acid binding moiety, a regulatory moiety. As used herein, "regulate" or "modulate" refers to an ability to alter the level of expression of a particular gene 20 above (i.e., up-regulate or activate) or below (i.e., down-regulate or repress) the basal level of expression that would occur in the particular system (for example, an in vitro transcription system or a cell) in the absence of the compound under the same conditions. A regulatory moiety that activates transcription is referred to herein as an "activation moiety" or "activator", whereas a regulatory moiety that 25 represses transcription is referred to as a "repressor moiety" or "repressor."

In general, a regulatory moiety is any compound that can positively or negatively effect, by either a direct mechanism (i.e., by direct interaction with one or more components of the transcription complex) or and indirect mechanism (i.e., by (i) direct interaction with a repressor protein or (ii) direct interaction with a protein 30 involved in chromatin or nucleosome structure), transcription of a target gene, other than by direct electrostatic interaction with double-stranded DNA. By "direct interaction" is meant direct, non-covalent association between two components.

Representative embodiments of regulatory moieties include peptides, polypeptides, lipids, carbohydrates, and any combination thereof. A "peptide" is a polymer (i.e., a linear chain of two or more identical or non-identical subunits joined by covalent bonds) made up of naturally occurring or synthetic D- or L-, or D- and L-, amino acids joined by peptide bonds. Generally, peptides contain at least two amino acid residues (i.e., the molecules resulting from the formation of a peptide bond between two amino acids, or between an amino acid residue and another amino acid) but fewer than about 50 amino acid residues. A "polypeptide" is also a polymer of amino acid residues linked by peptide bonds, but typically contains at least about 50 amino acid residues. Thus, herein "peptide" is used to refer to a regulatory moiety that is less than about 50 amino acid residues in length, and "polypeptide" refers to larger polymers of amino acid residues linked by peptide bonds. A "lipid" is a substantially water-insoluble molecule that contains as a major constituent an aliphatic hydrocarbon. Lipids include fatty acids, neutral fats, waxes, and steroids. The hydrocarbon portions of the molecule can be of any length, can be saturated or unsaturated, and can be straight- or branched-chain. "Carbohydrate" refers to any aldehyde or ketone derivative of a polyhydric alcohol, and includes starches, sugars, celluloses, and gums.

A regulator moiety can be naturally occurring or be derived from or an analog of a naturally occurring molecule. Alternatively, it can be synthetic. Preferred regulatory moieties are peptides and small organic molecules. "Peptide", in this context, refers to an amino acid polymer comprised of between one to about fifty, inclusive, residues of amino acids (i.e., any molecule that contains an amino group and a carboxylic group linked by peptide, or carboxamide, bonds. Such peptides can be duplicative of amino acid sequences that occur naturally (e.g., an activation or repressor domain of a transcriptional activator or repressor, respectively). Typically, such peptides are comprised of the twenty L amino acids commonly found in proteins in nature. Alternatively, they can comprise one or more D enantiomers of such amino acids, or other rarely observed amino acids. In other embodiments, such peptides are comprised of amino acids not typically found in naturally occurring proteins. Peptidic regulatory moieties can be synthesized by any suitable method, including recombinant techniques or solid state or in-solution

synthetic methods. After synthesis, they can be linked to a suitable nucleic acid binding moiety. If it is unknown whether a particular peptide possesses regulatory function, it can be screened in a suitable system. Indeed, large numbers of such peptides (as can, for example, result from combinatorially generating peptides) can
5 be screened in high throughput formats.

In certain preferred embodiments, the regulatory moieties are small organic molecules that have been identified as having regulatory function. By "small organic molecule" is meant any water soluble organic molecule having a molecular weight of less than about 10 kDa (kilo Dalton), preferably less than about 5 kDa,
10 more preferably less than about 2.5 kDa, even more preferably less than about 1.5 kDa, and even more preferably less than about 1 kDa.

In some embodiments, a synthetic regulatory compound comprises two, three, four, or more regulatory moieties linked to the nucleic acid binding moiety(ies). In such instances, the regulatory moieties can be the same or different,
15 and can serve to recruit or retard the same or different transcription factors.

Another component of the synthetic regulatory compounds of the invention is a linker, which can be any molecule that can be used to link at least one nucleic acid binding moiety to at least one regulatory moiety in a manner that allows the nucleic acid binding moiety(ies) to retain its(their) intended nucleic acid binding
20 function(s) and the regulatory moiety(ies) to retain its(their) ability to influence transcription of the target gene. The linker should provide adequate spacing between and/or orientation with respect to the nucleic acid binding moiety and the regulatory moiety so as to allow each to retain its respective function.

Linkers used in the practice of this invention can be branched, although
25 straight chain molecules are preferred. They can be amphipathic or aliphatic, with molecules in the latter class being preferred. Typically, a linker will contain from about one to about 200 "spacing" or "backbone" moieties (e.g., $-\text{CH}_2-$, $-\text{CH}=\text{}$, and $-\text{C}\equiv$). The backbone moieties preferably are heteroatoms including, but not limited to, carbon, nitrogen, oxygen, sulfur, and phosphorus, with the majority of such
30 atoms being carbon. One or more different side chain groups can also be appended to the backbone. Spacing between multiple side chain unit groups can be variable or

consistent along the backbone. Representative examples of suitable linkers include polyethylene glycol, alkyl chains, and peptides.

In particularly preferred embodiments of this aspect, the synthetic regulatory compounds are cell permeable. In other words, they are able to come into contact
5 with and be internalized by a cell. Such cells include animal (both prokaryotic and eukaryotic) and plant cells. Preferred animal cells include vertebrate cells, particularly arachnid, avian, fish, insect, and mammalian cells. Particularly preferred are mammalian cells, for example, bovine, avian, canine, equine, feline, human, murine, ovine, porcine, and primate cells. Preferred plant cells include those
10 of commercially important grains (e.g., alfalfa, barley, corn, rice, soy, sorghum, wheat) and ornamental plants.

Synthetic regulatory compounds of the invention can be prepared as pharmaceutical salts by any method known in the art depending upon the intended application.

15 A related aspect of the invention concerns compositions comprising a synthetic regulatory compound according to the invention and a carrier, for example, a pharmaceutically acceptable carrier. Such compositions can be dry or liquid formulations, and can optionally include one or more excipients, stabilizers, bulking agents, etc. Dry formulations include those that have been lyophilized or freeze
20 dried, which formulations can be reconstituted in a solvent prior to use. Liquid formulations include aqueous formulations, oils, emulsions, and suspensions.

Compositions according to the invention can be delivered to a subject by any suitable route. For example, in the context of compositions intended for agricultural use, a composition can be applied by spraying a liquid or broadcasting a solid. For
25 administration to animals, such compositions can be injected (e.g., via subcutaneous, intramuscular, intravenous, and other parenteral routes), inhaled, eaten, or delivered topically.

Another related aspect concerns kits containing a synthetic regulatory compound or composition according to the invention. Kits typically comprise one
30 or more synthetic regulatory compounds of the invention in a suitable composition and packaged in an appropriate storage container, for example, a vial or ampule. Kits often further comprise external packaging, such as a box or other container, to

protect and support the storage container containing the composition. In some embodiments, the packaging also contains directions for use, package inserts, etc.

Still another related aspect of the invention concerns complexes comprising a synthetic regulatory compound of the invention complexed with a nucleic acid molecule, e.g., a DNA, particularly a dsDNA. Such complexes can be formed, for example, by exposing a composition (for example, a cell or an in vitro transcription reaction) containing a nucleic acid (e.g., dsDNA) comprising a target nucleotide sequence to a synthetic regulatory compound according to the invention.

Thus, a related aspect concerns cells containing a synthetic regulatory compound of the invention. Such cells include animal cells and plant cells. Preferred animal cells include avian, bovine, canine, equine, feline, fish, human, murine, ovine, porcine, and primate cells. Such cells can be in vivo or in vitro (including ex vivo).

Another aspect of the invention relates to methods for regulating expression of a target gene by exposing the target gene and its associated regulatory elements to a synthetic regulatory compound according to the invention. Such methods can be carried out in vitro and in vivo.

Still another aspect of the invention concerns the use of a synthetic regulatory compound of the invention to prevent or treat a disease. Such methods can be accomplished by delivering an effective amount of the synthetic regulatory compound to an animal or plant. An "effective amount" refers to an amount of a synthetic regulatory compound sufficient to induce or effectuate a detectable change in transcription of a desired gene in a cell-free in vitro system or a cell, be it in an organism or in culture. What constitutes an effective amount will depend on a variety of factors that the skilled artisan will take into account in arriving at the desired delivery regimen.

Another aspect of the invention concerns methods of screening for synthetic regulatory compounds from amongst one or more test compounds. These screening methods include both in vitro and in vivo screening methods, and can include methods involving an in vitro screen followed by an in vivo screen (e.g., a cell-based screen). Such methods can be performed, for example, by exposing, under transcription conditions, a dsDNA encoding a regulatable gene to a test compound

comprising a nucleic acid binding moiety targeted to a transcription-associated regulatory element of the regulatable gene conjugated via a linker to a regulatory moiety, and determining whether the test compound regulates expression of the regulatable gene. Preferably, such methods are performed in vitro, preferably in a high throughput format, meaning that more than about 10, preferably, more than
5 about 100, 1,000, or 10,000 compounds are screened at once. Preferably, the regulatable gene is a marker gene, such as a gene encoding a luciferase or green fluorescent protein.

The various moieties of the invention can be synthetic or natural products.
10 Synthetic moieties can be synthesized by solution or solid phase methods. Two or more moieties can also be synthesized together. Compounds according to the invention can be in unpurified, substantially purified, and purified forms. The compounds can be present with any additional component(s) such as a solvent, reactant, or by-product that is present during compound synthesis or purification,
15 and any additional component(s) that is present during the use or manufacture of a compound or that is added during formulation or compounding of a compound.

Another aspect relates to methods for synthesizing the compounds of the invention. Broadly, such methods typically comprise separately obtaining each of the nucleic acid binding moiety, the linker, and the regulatory moiety to be linked
20 together to form the synthetic regulatory compound. Two or more of these moieties are then linked by a suitable chemistry, followed by linkage to the third moiety. The particular order of addition of moieties can vary, and its selection is within the skill of the ordinary artisan. Alternatively, the linker and regulatory moiety, or linker and nucleic acid binding moiety, can be synthesized as a single unit. In other
25 embodiments, all three elements are synthesized together. Following synthesis, the compound, or its intermediary compounds in intermediary steps, is preferably purified.

The above summary of the invention is not limiting and other features and advantages of the invention will be apparent from the following Brief Description of
30 the Figures and Detailed Description, as well as from the appended Claims and Abstract.

BRIEF DESCRIPTION OF THE FIGURES

SUBSTITUTE SHEET (RULE 26)

Figure 1 illustrates a double-stranded DNA (dsDNA) molecule comprising an inverted tandem repeat of the target nucleotide sequence of an eight ring hairpin polyamide molecule (SEQ ID NO: 1). In the polyamide portion of the synthetic regulatory compound, imidazole moieties are represented by blackened circles and pyrroles are represented by open circles. Diamonds represent β -alanine. Amino acids are represented by the standard one letter code. This nomenclature is used consistently throughout the specification unless otherwise expressly indicated. Two synthetic regulatory compounds, each comprised of the same components, are illustrated as being bound to their respective target sites in the dsDNA. Linked to each polyamide through a linker is one of three activator peptides. The compounds are designated 7 (SEQ ID NO: 2), 8, and 9 (SEQ ID NO: 3). In compound 7, the linker comprises of a Cys residue linked to an XL peptide. In compound 8, the linker is larger and contains a Cys residue at its amino-terminus, and an XL peptide at its carboxy-terminus. Compound 9 comprises a 28 amino acid residue peptide, additionally including a Cys residue at the N-terminus and a XL peptide at its C-terminus.

Figure 2 contains 3 panels, A, B, and C. Panels A and B show dsDNA comprising inverted tandem repeats of target nucleotide sequences for a specific polyamide. In panel A, the polyamide has an activating peptide (SEQ ID NOs: 4 and 5) attached to its C-terminus via a linker. In panel B, the polyamide has the linker-activating peptide conjugate attached via the C-terminal pyrrole residue. In such circumstances, the polyamide contains conventional C-terminal moieties, for example, β -Dp. Panel C depicts a single hairpin polyamide to which two linker-activating peptide domains are attached via internal pyrrole moieties and SEQ ID NOs: 6 and 7). Also described are compounds 10-20, specifically the portions of the compounds comprising the linker and activating domains.

Figure 3 is a graphic representation of the template used for *in vitro* transcription assays. As shown, the template contains three inverted tandem repeat target nucleotide sequences (hatched region) to be targeted by nucleic acid binding moiety of a synthetic regulatory compound. The 3'-terminus of the downstream most of these sequences is 50 based pairs upstream of the adenovirus major late

promoter TATA box that drives the expression of a “g-less” transcript. In each inverted tandem repeat sequence, the number of nucleotides between the target nucleotide sequences is 3, 5, or 7.

Figure 4 depicts a graphic illustration of a double stranded nucleic acid molecule (SEQ ID NOS: 8 and 9) containing an inverted tandem repeat target nucleotide sequence for an eight ring polyamide which provides DNA sequence target specificity. Also depicted are the chemical structures of N-methylpyrrole and N-methylimidazole sub-units that may be incorporated into nucleic acid targeting moieties, *e.g.*, polyamides, for use in conjunction with the present invention. Also shown are two representative linker molecules that may be used to join nucleic acid binding moiety and a regulatory moiety in order to generate a synthetic regulatory compound according to the invention. The figure also lists the compounds that were generated upon the ligation of the respective peptides to the polyamide.

Figure 5 depicts synthesis of synthetic activators. The synthesis of polyamide 1 was accomplished according to established protocols. (a) Treatment of compound 1 with thiolane-2,5-dione followed by benzyl bromide provided thioester compound 2 in good yield (53%). (b) Combination of thioester-containing compound 2 with each peptide (SEQ ID NOS: 10 and 11) in denaturing buffer then provided the targeted conjugates via the native ligation reaction.

Figure 6 depicts substitution of the dimerization module with a flexible ethylene glycol-derived linker. The synthesis of hairpin polyamides 6 and 7 and compounds 8 and 9 was carried out as described.

Figure 7 illustrates a synthetic regulatory compound comprising a hairpin polyamide (linked head to tail by a short linker (*e.g.*, γ -aminobutyric acid)) attached to a linker domain (LD) to which an activation domain (AD) is attached. The polyamide is associated with the minor groove of a DNA double helix via its DNA recognition domain. Also illustrated is a dsDNA comprising a target nucleotide sequence with which is associated a hairpin polyamide targeted to that sequence. The chemical structure of pyrrole moieties (open circles) are shown, as is the structure of an imidazole residue (shaded circle).

Figure 8 depicts compound 3 (PA-Gcn4-AH) bound to its cognate palindromic DNA site and activated transcription in vitro when its predetermined

DNA binding sites are present (SEQ ID NOs: 12 and 13). (A) (Upper) Storage phosphor autoradiogram of a quantitative DNase I footprinting titration of compound 3 on the 3-³²P-labeled 271-bp pPT7 *EcoRI/PvuII* restriction fragment carried out according to established protocols. Pre-equilibration of compound 3 with the DNA fragment was carried out for 75 min., before initiation of the cleavage reactions. From left to right the lanes are: the A sequencing lane; DNase I digestion products in the presence of compound 3 at concentrations of 100 nM, 50 nM, 25 nM, 10 nM, 5 nM, 2.5 nM, 1 nM, 0.5 nM, and 0.25 nM, respectively; DNase I digestion products with no compound 3 present; undigested DNA. (Lower) Data for compound 3 in complex with the 19-bp palindromic site. The curve through the data points is the best-fit cooperative Langmuir binding titration isotherm (n_2) obtained from a nonlinear least-squares algorithm. (B) An in vitro transcription reaction containing PA-Gcn4-AH (compound 3) at 200 nM shows enhanced expression of a 277-nt transcript relative to basal levels whereas a reaction containing conjugate 4, lacking the activating region, does not. Inclusion of polyamide 1 (lane 2) in the reaction did not impair basal transcription (lane 1) match template (SEQ ID NO: 14), mismatch template (SEQ ID NO: 15). The variation in transcript position for lane 4 was found to be caused by curvature of the gel. (C) In vitro transcription reactions containing 3 (PA-Gcn4-AH) with templates bearing either the cognate palindromic binding sites (match template) or palindromic sites in which a G:C base pair has replaced a T:A base pair in each half site (mismatch template) upstream of the core promoter. The concentrations of compound 3 used were 0 (basal), 10 nM, 100 nM, and 500 nM.

Figure 9 Depicts dependence of activation level upon time and activating region. (A) (Upper) Storage phosphor autoradiogram showing in vitro transcription time course experiment with compounds 3 and 4 present at 300 nM concentration. Aliquots were processed at 10, 20, 30, 40, and 60 min. (Lower) Comparison of the amount of transcript obtained at each time point relative to basal transcription levels in which no conjugate was present (fold activation). (B) (Upper) Storage phosphor autoradiogram showing the effect of increasing concentrations of untethered 21-aa AH peptide on transcription reactions containing either compounds 4 or 3 at 300 nM concentration. Transcription reactions were performed for 30 min, in the presence

SUBSTITUTE SHEET (RULE 26)

of 0, 0.2 :M, 2 :M, and 10 :M concentrations of AH peptide. (Lower) For each reaction, the amount of transcript obtained was compared with basal transcription levels to give the respective fold activation value. These values are displayed as percent activation compared with the results from the reaction containing compound 3 (lane 5), which is defined as 100%.

Figure 10 depicts the activation levels for compounds 3, 5, 8, and 9 were determined by comparison with the amount of transcript obtained from reactions containing the relevant parent hairpin polyamides. A) The fold activation values thus obtained are displayed as percentages relative to the fold activation mediated by compound 3, defined as 100%. (B) Data from DNase I footprinting titrations with compounds 5 and 8. The curve through each data set is the best-fit Langmuir binding titration isotherm ($n = 1$) obtained from a nonlinear least-squares algorithm.

Figure 11 describes the structure of three polyamides, 1, 5, 9. With respect to compounds 1 and 5, they are separately derivatized with either of three peptides, designated AH, VP2 (SEQ ID NO: 16), and VP1 (SEQ ID NO: 17), the amino acid sequences of which are shown and the corresponding conjugate is numbered. In contrast, compound 9 shows a hairpin polyamide linked to a VP2 peptide through a linker attached to the C-terminal-most pyrrole of the polyamide.

Figure 12 shows the ability of various conjugates to activate transcription in vitro. Background levels of transcription in the absence of any polyamides are shown in lane 1. Addition of the polyamide alone does not alter the levels of transcription appreciably (lane 2). Lane 3 shows the levels of transcription elicited by conjugate 2; lane 4 shows activation elicited by conjugate 4 and lane 5 demonstrates the activity of conjugate 3. Lane 6 shows activation elicited by conjugate 9. Lane 7 represents the activation due to conjugate 6; lane 8 corresponds to conjugate 8 and lane 9 is due to conjugate 7. Reactions in lanes 1-9 were performed with on a template bearing three cognate palindromic sites (first described in Figure 4) whereas lanes 10-18 were performed in the same order as 1-9 except a template bearing non-cognate sites upstream of the promoter was used. In each of the reactions 400nM concentrations of the relevant conjugate were used.

Figure 13 shows the structures of compounds 10 and 11, as well as a dsDNA molecule (SEQ ID NOs: 18 and 19) containing a target nucleotide sequence for the

SUBSTITUTE SHEET (RULE 26)

polyamide, wherein the polyamide is attached via the PEG linker (shaded hexagon) to the VP2 regulatory peptide. These conjugates recognize a different dsDNA sequence.

Figure 14 shows the modular nature of the synthetic regulator. In vitro transcription reactions show that substituting the imidazole polyamide (conjugate 1 of Figure 11) for that described in Figure 13 (conjugate 10) alters the template recognition properties in a specific manner. Lanes 1-7 represent reactions performed with the template designed to bind polyamide 10 and its various conjugates (#11) and lanes 8-14 show reactions which were performed with template 1 described in Figure 4. Lanes 2 & 3 using template 2 and lanes 9&10 on template 1 show the inability of either polyamide (at 400nM) to influence transcription levels. Lanes 4&5 on template 2 and lanes 13 &14 on template 1 show that the corresponding polyamide conjugate elicits transcription on the appropriate template. Lanes 6&7 for conjugate 7 and lanes 11& 12 for conjugate 11 show that on the non-conjugate sites the regulatory moiety bearing conjugates do not elicit transcription above background levels.

DETAILED DESCRIPTION OF THE INVENTION

The synthetic regulatory compounds of the invention represent a novel class of nucleic acid binding ligands in which a nucleic acid binding moiety is tethered through a linker to a second functional moiety, viz, a regulatory moiety (e.g., an activator), that interacts directly with elements of the endogenous transcriptional machinery or indirectly through interaction with a repressor protein or components involved in chromatin structure (e.g., nucleosomes). Cell-permeable members of this class can be targeted to designated sites in the genome, and can be used, for example, as tools to study mechanistic aspects of transcriptional regulation and to correct the ectopic gene expression that often occurs in disease.

The invention now will be discussed with reference to particular preferred embodiments, that, for convenience, will be in the context of dsDNA as the nucleic acid, but it is to be understood that the invention is not limited to such context and can be applicable to other nucleic acid, i.e., single-stranded DNA or single- or double-stranded ribonucleic acid.

Synthetic Regulatory Compounds

Synthetic regulatory compounds of this invention are now discussed in greater detail, especially with reference to preferred embodiments thereof.

1. Nucleic Acid Binding Moieties.

One component of the synthetic regulatory compounds of the invention
5 concerns nucleic acid binding moieties. In the context of this invention, a nucleic acid binding moiety is any compound or chemical that can provide site-specific recognition of a target nucleotide sequence in a nucleic acid molecule preferably a dsDNA. In preferred embodiments, such specificity is provided by an ability to recognize specific based pairs within the major groove and/or minor groove of a
10 double-stranded nucleic acid molecule. Interactions of this sort are typically mediated by electrostatic forces, hydrogen bonding, van der Waals forces, and steric considerations. In natural systems, such binding specificity is typically mediated by peptides that are comprised within protein-based transcription factors and other proteins that bind to nucleic acids within cells and viruses. Herein, such peptides are
15 referred to as "natural DNA binding ligands," and are not within the scope of the instant invention as it relates to embodiments that comprise but one nucleic acid binding moiety. However, in embodiments that comprise two or more nucleic acid binding moieties, one or more such natural peptides can be included in the synthetic regulatory compound, provided that at least one of such nucleic acid binding
20 moieties is a non-natural nucleic acid binding moiety. Thus, any such peptides, whether now known or later developed, can be included in synthetic regulatory compounds that comprise a plurality of nucleic acid binding elements.

Below, various preferred embodiments of nucleic acid binding moieties useful in the practice of the present invention are provided.

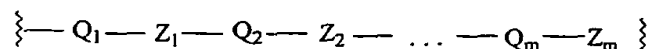
25 A. Molecular Scaffolds

In certain preferred embodiments of the invention, the nucleic acid binding moiety comprises a molecular scaffold. A molecular scaffold is any compound that spatially arises a plurality of hydrogen bond donors and acceptors in a manner that allows the formation of hydrogen bonds with the corresponding hydrogen bond
30 acceptors and donors in a target nucleotide sequence in a nucleic acid molecule. The structures of single- and double-stranded nucleic acid molecules at high resolution, such as the three-dimensional coordinates of each individual atom within the nucleic

acid molecule are known to a resolution of less than about 2 Å. Such structures can be determined by techniques known in the art, for example, x-ray crystallography or NMR spectroscopy. Accordingly, a high-resolution model of any particular nucleotide sequence within a nucleic acid molecule can be determined and used for modeling purposes. From such models, it is possible to deduce the positions of various hydrogen bond donor and acceptor elements within a nucleic acid molecule. The distance ranges over which such interactions occur are also known in the art, as are such disinteractions for other forces involved in nucleic acid molecule/nucleic acid binding ligand interactions.

From such data, it is possible to design synthetic ligands that correspondingly array hydrogen bond acceptors and donors in a manner that allows formation of hydrogen bonding upon interaction of the particular ligand with a particular nucleotide sequence in such nucleic acid molecule. This invention encompasses designing a molecular scaffold directed toward a particular target nucleotide sequence. Such molecules can then be screened in vitro and/or in vivo to ascertain if the desired interaction occurs. Those molecules exhibiting the desired interaction, particularly those that do so at submacromolar and preferably subnanomolar binding affinities can be selected for further use. Alternatively, such compounds can be used as a basis for lead optimization in order to generate additional analogs that exhibit the desired interactions. Preferably, any such molecule selected for further use will also exhibit cell-permeability.

In yet other embodiments, the nucleic acid binding moiety comprises the structure



where each of Q_1, Q_2, \dots, Q_m is a heteroaromatic moiety or $(\text{CH}_2)_p$ (where p is an integer between 1 and 3, inclusive); each of Z_1, Z_2, \dots, Z_m is a covalent bond or a linking group; and m is an integer between 1 and 9 (preferably between 2 and 4), inclusive. Where Q is a heteroaromatic moiety, it is preferably selected from optionally substituted imidazole, pyrrole, pyrazole, furan, isothiazole, oxazole, isoxazole, thiazole, thiophene, furazan, 1,2,3-thiadiazole, 1,2,4-thiadiazole, 1,2,5-thiadiazole, 1,3,4-thiadiazole, 1,2,3-triazole, 1,2,4-triazole, 1,3,4-oxadiazole, 1,2,4-

oxadiazole, and thiophene moieties. Exemplary substituents include Cl, F, CH₃ (e.g., as in N-methylpyrrole or N-methylimidazole), hydroxy (e.g., as in 3-hydroxypyrrole).

Linking groups Z₁, Z₂, ... Z_m, can be between 2 and 5 (preferably 2) backbone atoms long. Exemplary linking groups include carboxamide, amidine, and ester groups, with carboxamide groups being preferred.

Any suitable synthetic method can be used to link the various elements of the compounds of this invention. In addition to those described in the examples further below, other suitable methods now known in the art or later developed can be used.

10 B. PNAs

Protein nucleic acids (PNAs) are analogs of DNA in which the backbone is structurally homomorphous with a deoxyribose backbone, except that the backbone linkages are peptide bonds rather than phosphate esters. The PNA backbone comprises N-(2-aminoethyl)glycine units to which the nucleobases are attached.

15 PNAs containing all four natural nucleobases hybridize to complementary oligonucleotides according to Watson-Crick base-pairing rules, and thus represent true DNA mimics in terms of base-pair recognition. Egholm et al. (1993) Nature 65:566-568. Since a PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes exhibit greater thermal stability, as compared to DNA/DNA, DNA/RNA, or RNA/RNA duplexes. PNAs have the additional advantage in not being
20 recognized by nucleases or proteases. In addition, PNAs can be synthesized on an automated solid state synthesizer using standard t-Boc chemistry. The design and synthesis of PNA molecules are described, for example, in U.S. Patent Nos. 5,539,083; 5,864,010; 5,977,296; and 5,985,563.

25 PNA-based non-natural nucleic acid binding moieties can be designed, synthesized, and incorporated into compounds according to the invention. Chemistries suitable for attaching linkers (or linkers already conjugated to one or regulatory moieties) to such PNAs are known in the art and can be used for such purposes.

30 C. Triplex-Forming Oligonucleotides

Other preferred embodiments of the invention concern synthetic regulatory compounds wherein the nucleic acid binding moiety is an oligonucleotide capable of

base-pair specific recognition with a double-stranded nucleic acid, such that the oligonucleotide can form a triple helix. Oligonucleotide-directed triple helix formation is one of the most effective methods for accomplishing the sequence specific recognition of double helical DNA. See, e.g., U.S. Patent No. 5,847,555; Moser et al. (1987) Science 238:645; Le Doan et al. (1987) Nucl. Acids Res. 15:7749; Maher et al. (1989) Science 245:725; Beal et al. (1991) Science 251:1360; Strobel et al. (1991) Science 254:1639; and Maher et al. (1992) Biochem. 31:70. Triple helices form as the result of hydrogen bonding between bases in a third strand of DNA and duplex base pairs in the double stranded DNA, via Hoogsteen base pairs.

Briefly, such oligonucleotides typically comprise between about 10 to about 200, preferably about 10 to about 50, nucleotides. Such oligonucleotides can be synthesized by any suitable method, for example, by conventional automated solid state techniques. Typically, nucleotide sequences targeted by triplex-forming oligonucleotides comprise purine rich tracts on one of the strands of the double-stranded, double-helical nucleic acid. The triple helix so formed contains the oligonucleotide bound in either a parallel or anti-parallel orientation with respect to the target sequence depending on the nucleotide sequences used in the oligonucleotide.

A parallel orientation occurs when the oligonucleotide is a pyrimidine-rich oligonucleotide. In particular, the pyrimidine-rich oligonucleotide contains a thymine containing nucleotide (T) when the nucleotide at the complementary position in the purine-rich target sequence is an adenosine containing nucleotide (A) and a cytosine containing nucleotide (C) when the nucleotide at the complementary position in the purine-rich target sequence is a guanine containing nucleotide (G).

An anti-parallel orientation occurs when a purine-rich oligonucleotide is used. In particular, anti-parallel orientation is obtained when the purine-rich oligonucleotide contains a guanine containing nucleotide (G) when the nucleotide at the complementary position in the purine-rich target sequence is a guanine containing nucleotide (G) and an adenosine containing nucleotide (A) when the nucleotide at the complementary position in the purine-rich target sequence is an adenosine containing nucleotide (A).

SUBSTITUTE SHEET (RULE 26)

Synthetic triple helix-forming oligonucleotides can also bind in an anti-parallel orientation to a purine-rich target sequence. Such triple helix-forming oligonucleotides contain a G when the nucleotide in the complementary position of the purine-rich target sequence is G, T or A when the nucleotide in the complementary position in the purine-rich target sequence is an \bar{A} and the nucleotide nebularine when the complementary position in the purine-rich target sequence is C.

Because such moieties are capable of targeting specific nucleotide sequences in double-helical nucleic acids, they represent suitable nucleic acid moieties for use in developing candidate synthetic regulatory compounds.

Preferably, such oligonucleotides (as with other nucleic acid binding moieties of the invention) will be designed to target a sequence that will allow a regulatory moiety attached thereto through a linker to be brought into proximity of the promoter of the gene, the expression of which is desired to be regulated. Using the nucleotide sequence of the region proximal to the target gene's promoter, one or more oligonucleotides can be designed that will possess the desired triplex-forming function.

The oligonucleotides used in the invention to form triple helices can be made synthetically by well-known synthetic techniques to contain a structure corresponding to the naturally occurring polyribonucleic or polydeoxyribonucleic acids. See, e.g., U.S. Patent No. 5,847,555. Alternatively, the phosphoribose backbone of such oligonucleotides can be modified such that the thus formed oligonucleotide has greater chemical and/or biological stability. Biological stability of the oligonucleotide is desirable when the oligonucleotides are used in vivo. Such modified oligonucleotides can be synthesized with a structure that is stable under physiological conditions, including enhanced resistance to nuclease degradation. Further, when used in vivo, such nucleotides preferably have a minimal length that permits targeted triple helix formation so as to facilitate the transport of the oligonucleotide across the membranes of the cytoplasm and nucleus.

These and other nucleic acid binding moieties intended for use in a synthetic regulatory compound of the invention are preferably tested against the target nucleotide sequence to identify those which bind with the highest affinity and greatest specificity. Binding affinity can be assessed by any suitable method, for

SUBSTITUTE SHEET (RULE 26)

example, DNase I footprinting. Specificity can be determined by using nucleic acids containing target sequences and others containing nucleotide sequences that differ from the target sequence by one or more nucleotides. Preferably, such compounds have binding affinities (in terms of association constants) of at least about 10^6 M^{-1} , more preferably at least about 10^9 M^{-1} , and even more preferably, at least about 10^{10} M^{-1} , 10^{11} M^{-1} , 10^{12} M^{-1} , or higher. Specificities of least about 2, preferably at least about 3, 4, or 5, even more preferably at least about 10, 50, 100, or more, for target versus single base pair "mismatch" sites are preferred.

Such molecules can then serve as the basis for the synthesis of the synthetic regulatory compounds according to the invention by attaching one or more regulatory moieties thereto through one or more linkers. The resulting compounds are then preferably tested in cell-based assays (e.g., where the cells carry a reporter construct that comprises a reporter gene under the control of a promoter region that is targeted by the nucleic acid binding moiety of the compound) to determine if the desired regulatory function is achieved by the particular synthetic regulatory compound. If so, the compound typically will represent a lead compound which can then be used to develop potential therapeutic or prophylactic drugs or other compounds that can be administered to cells to achieve the desired regulatory function.

20 D. Polyamides

As described above, the nucleic acid binding moiety includes those that are dsDNA intercalators, dsDNA minor groove binding moieties, and dsDNA major groove binding moieties. It is to be understood that, where the nucleic acid binding moiety is referred to as a "minor groove binder" (or words to that effect), it does not mean that such moiety has binding interactions exclusively with the minor groove; the moiety also can have binding interactions with other parts of the dsDNA, for example, with adjacent base pairs by intercalation, with backbone phosphate groups, or with the major groove.

In certain preferred embodiments, the nucleic acid binding moiety is a minor groove binder, which typically (but not necessarily) has an elongate crescent shape, topologically complementary to the shape of the minor groove. The minor groove binder can be a residue of a naturally-occurring compound, such as doxorubicin,

daunomycin, anthramycin, calicheamycin, mitomycin, duocarmycin, distamycin, and netropsin, or an analog or a derivative thereof. Alternatively, the nucleic acid binding moiety can be a residue of a synthetic minor groove binder.

In particularly preferred embodiments, the nucleic acid binding moiety is a synthetic polyamide unit comprising N-methylpyrrole carboxamide ("Py") units and optionally one or more of N-methylimidazole carboxamide ("Im"), N-methyl-3-hydroxypyrrole carboxamide ("Hp"), glycine carboxamide, P-alanine carboxamide, γ -aminobutyric acid carboxamide, 5-aminovaleric acid carboxamide, and γ -2,4-diaminobutyric acid carboxamide units.

Such synthetic polyamides are minor groove binders, binding with high binding constants, often greater than 10^9 M^{-1} . The design and synthesis of such polyamides are described for instance in Baird et al. (1996) J. Am. Chem. Soc. 118:6141-6146; and U.S. Applications 08/607,078 (filed Feb. 26, 1996); 09/374,702 (filed Aug. 12, 1999); 09/372,473 (filed Aug. 11, 1999); 09/372,474 (filed Aug. 11, 1999); 09/414,611 (filed Oct. 8, 1999); and 60/115,232 (filed Jan. 6, 1999, the benefit and priority of which is claimed by international application PCT/US00/00298, entitled "Compositions and Methods Relating to Cyclic Compounds that Undergo Nucleotide Base Pair Specific Interactions with Double Stranded Nucleic Acids"), the disclosures of which are incorporated herein by reference. It has been further discovered that such polyamides can bind to dsDNA with two heteroaromatic carboxamide moieties fitting side-by-side within the minor groove and that such side-by-side heteroaromatic carboxamide pairs recognize specific dsDNA base pairs, giving rise to a set of "pairing rules" correlating heteroaromatic carboxamide pairs and the DNA base pairs recognized:

<u>Heteroaromatic Pair</u>	<u>dsDNA Base Pair(s) Recognized</u>
Im/Py	G/C
Py/Im	C/G
Py/Py	A/T, T/A
Hp/Py	T/A
Py/Hp	A/T

Where it is desired to synthesize a polyamide-based nucleic acid binding moiety that binds to dsDNA with specificity for particular base pair sequences, resort can be had to the above pairing rules.

Glycyl or β -alanyl carboxamides can serve as "spacer" groups for adjusting
5 the position of the heteroaromatic carboxamide residues in relation to the nucleotide
base pairs of a polyamide-based nucleic acid binding moiety's binding site. A γ -
aminobutyric acid carboxamide, 5-aminovaleric acid carboxamide, or γ -2,4-
diaminobutyric acid carboxamide unit (or other moieties that produce a substantially
equivalent structural effect) provides the potential for formation of a "hairpin"
10 conformation, wherein some or all of the heteroaromatic carboxamide units from a
portion of the polyamide to one side of the turn-providing moiety bind side-by-side
to the heteroaromatic carboxamide units from the portion the polyamide to the other
side of the turn-providing moiety. See Figure 2C for a representative diagrammatic
illustration of polyamide hairpin conformation. Hairpin polyamides are capable of
15 targeting predetermined DNA sequences with affinities and specificities comparable
to DNA binding proteins in accordance with a simple set of pairing rules dictated by
the side-by-side binding of the aromatic amino acids. Mrksich et al. (1992) Proc.
Natl. Acad. Sci. USA 89:7586-7590; Wade et al. (1992) J. Am. Chem. Soc.
114:8784-8794; and Trauger et al. (1996) Nature 382:559-561. These synthetic
20 DNA binding ligands are cell permeable, and one such compound was shown to
specifically interfere with gene expression in mammalian cell culture. Gottesfeld et
al. (1997) Nature 387:202-205; and Dickinson et al. (1998) Proc. Natl. Acad. Sci.
USA 95:12890-12895.

When fewer than all of the heteroaromatic carboxamide units from one end
25 of the polyamide associate side-by-side with heteroaromatic carboxamide units from
the other side of the polyamide, the unpaired heteroaromatic carboxamide units from
one polyamide can be available to form cooperative side-by-side pairings with
unpaired heteroaromatic carboxamide units from another polyamide, for example
another hairpin or straight-chain polyamide. Such cooperative interaction can serve
30 to increase DNA binding specificity. Use of two turn-providing (or other non-
recognition moieties), for example, at each end of the nucleic acid binding moiety or

at one end and at an internal position, allows the formation of nucleic acid binding moieties having other conformations (e.g., cyclic or "H-pin" conformations, respectively). The 2-amino group of γ -2,4-diaminobutyric acid provides, among other locations, an attachment point for tandem-linked polyamide units, as well as
5 providing a moiety that can be used to introduce chirality into the polyamide. A Py, Hp, or Py equivalent heteroaromatic carboxamide can be replaced with a β carboxamide to form pairs such as β/β , β/Py , or Py/β . These and other molecular design principles disclosed in the aforementioned references can be used in the design of preferred examples of polyamide-based nucleic acid binding moieties of
10 the synthetic regulatory compounds of this invention.

Compounds of this invention are useful because they are strong nucleic acid binders, often as nanobinders (i.e., association constant (K_a) of 10^9 M^{-1}) or even as
picobinders (K_a of 10^{12} M^{-1}). It is especially noteworthy that some compounds of the invention are nanobinders while having relatively few heteroaromatic moieties
15 (3-5), while previously described nanobinders have generally required a larger number of heteroaromatic moieties.

Additionally, compounds of this invention will have anti-fungal (e.g., yeast, filamentous fungi) and/or anti-bacterial (Gram-positive, Gram-negative, aerobic, anaerobic) properties and therefore can be used for combating (i.e., preventing
20 and/or treating) infections by such pathogens. Other pathogens against which compounds of this invention can be used include protozoa and viruses. For human anti-infective applications, a compound of this invention can be used in combination with a pharmaceutically acceptable carrier. The composition can be dry, or it can be a solution. Treatment can be reactive, for example, to combat an existing infection,
25 or prophylactic, for preventing infection in an organism susceptible to infection.

Host organisms that can be treated include eukaryotic organisms, in particular plants and animals. The plant can be an agriculturally important crop, such as wheat, rice, corn, soybean, sorghum, and alfalfa. Animals of interest include mammals such as bovines, canine, equines, felines, ovines, porcines, and primates
30 (including humans).

While not wishing to be bound by any particular theory, it is believed that the synthetic regulatory compounds of this invention derive their biological activity by binding to double stranded nucleic acid, in particular double stranded DNA, and recruiting (in the case of activators) transcriptional machinery to the gene to be expressed or, in the case of repressors, inhibiting such recruitment.

The matching of a synthetic regulatory compound of this invention against a particular gene can be accomplished by rational design if the desired target dsDNA base pair sequence - e.g., a sequence in a gene (particularly a regulatory region thereof, e.g., a promoter, enhancer) is known. In such circumstances, a nucleic acid binding moiety that binds to the target base pair sequence with the desired degree of specificity is preferably used. The NABM can be a residue of a naturally occurring dsDNA binder with known specificity for the target sequence, or can be a synthetic dsDNA binder synthesized according to the base pair recognition rules discussed hereinabove. Alternatively, the matching can be accomplished by a suitable screening method.

i. Polyamide Synthesis.

There are two basic methods for synthesizing peptides and polyamides: the chemistry is either carried out in solution (solution phase) or on a solid support (solid phase). A major disadvantage of solution phase synthesis is the poor solubility of protected intermediates in organic solvents. Additionally, solution phase synthesis requires difficult purification methods. Solid phase synthesis avoids these problems, and thus is the preferred method in synthesizing peptides and polyamides.

U.S. Patent Nos. 6,090,947 and 5,998,140 detail solid state synthetic processes for use in synthesizing polyamides useful in the practice of this invention. Briefly, to make polyamides, such methods involve providing a solid support, preferably a polystyrene resin, for the stepwise addition of amino acids. To begin, the appropriate amino acid monomer or dimer is protected at its amino (NH₂) group with a Boc-group or an Fmoc-group and activated at the carboxylic acid (COOH) group by formation of an -OBt ester. The protected and activated amino acids are then sequentially added to the solid support, beginning with the carboxy terminal

SUBSTITUTE SHEET (RULE 26)

amino acid. When the desired polyamide has been prepared, the amino acids are deprotected and the peptide is cleaved from the resin and purified.

See also application no. PCT/US97/12722, filed 7/21/97; application no. PCT/US00/00298, filed 01/06/00, publication no. WO 00/40605, publication date 07/13/00; application no. PCT/US98/01714, filed 01/29/98, publication no. WO 98/37067, publication date 08/27/98; application no. 09/372,474, filed 08/11/99; application no. PCT/US98/06997, filed 04/08/98, publication no. WO 98/49142, publication date 11/05/98; application no. 08/837,524, filed 04/21/97; application no. 09/181,306, filed 10/28/98; application no. 09/374,702, filed 08/12/99; application no. PCT/US 98/03829, filed 01/29/98, publication no. WO 98/45284, publication date 10/15/98; application no. 08/853,522, filed 05/08/97; application no. 09/372,473, filed 08/11/99; application no. PCT/US98/01006, filed 01/21/98, publication no. WO 98/37066, publication date 08/27/98; application no. PCT/US98/02684, filed 02/13/98, publication no. WO 98/37087, publication date 08/27/98; application no. 09/374,704, filed 08/12/99; application no. 8/853,525, filed 05/08/97; application no. 09/367,513, filed 02/11/98; application no. PCT/US98/02444, filed 02/11/98; application no. 09/434,290, filed 11/05/99; application no. 09/359,921, filed 07/22/99; application no. 09/360,840, filed 07/22/99; application no. 09/360,344, filed 07/22/99; application no. 09/360,345, filed 07/22/99; application no. PCT/US99/20971, filed 09/10/99, publication no. WO 00/15242, publication date 03/23/00; application no. PCT/US99/20489, filed 09/10/99, publication no. WO 00/15773, publication date 03/23/00; application no. 09/414,611, filed 10/08/99; application no. 60/161,545, filed 10/26/99; application no. 09/479,279, filed 1/6/00; application no. 60/178,821, filed 01/28/00.

25 ii. Exemplary protocol for DNaseI footprint titration experiments

 All reactions were executed in a total volume of 400 μ L. A polyamide stock solution or H₂O (for reference lanes) was added to an assay buffer containing 3'-³²P radiolabeled restriction fragment (20,000 cpm), affording final solution conditions of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl₂, 5 mM CaCl₂, pH 7.0. The solutions were allowed to equilibrate for at least 12 hours at 22°C. Footprinting reactions were initiated by the addition of 10 μ L of a stock solution of DNase I (at the

appropriate concentration to give ~55% intact DNA) containing 1 mM dithiothreitol and allowed to proceed for 7 minutes at 22°C. The reactions were stopped by the addition of 50 µL of a solution containing 2.25 M NaCl, 150 mM EDTA, 23 µM base pair calf thymus DNA, and 0.6 mg/ml glycogen, and ethanol precipitated. The reactions were resuspended in 1 x TBE/ 80% formamide loading buffer, denatured by heating at 85°C for 15 minutes, and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% crosslinking, 7 M urea) in 1 x TBE at 2000 V for 1.5 h. Gels were dried on a slab dryer and then exposed to a storage phosphor screen at 22°C.

10 iii. Quantitative DNase I Footprint Titration Data Analysis.

Background-corrected volume integration of rectangles encompassing the footprint sites and a reference site at which DNase I reactivity was invariant across the titration generated values for the site intensities (I_{site}) and the reference intensity (I_{ref}). The apparent fractional occupancy (θ_{app}) of the sites were calculated using the equation:

$$\theta_{app} = 1 - \frac{I_{site} // ref}{I_{site}^{\circ} // ref^{\circ}} \quad (1)$$

where I_{site}° and I_{ref}° are the site and reference intensities, respectively, from a DNase I control lane to which no polyamide was added.

The ($[L]_{tot}$, θ_{app}) data were fit to a Langmuir binding isotherm (eq. 2, $n=1$) by minimizing the difference between θ_{app} and θ_{fit} , using the modified Hill equation:

$$\theta_{fit} = \theta_{min} + (\theta_{max} - \theta_{min}) \frac{K_a [L]_{tot}}{1 + K_a [L]_{tot}} \quad (2)$$

where $[L]_{tot}$ is the total polyamide concentration, K_a is the equilibrium association constant, and θ_{min} and θ_{max} , are the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (v. 3.0.1,

Abelbeck Software) with K_a , θ_{\max} , and θ_{\min} , as the adjustable parameters. The goodness of fit of the binding curve to the data points is evaluated by the correlation coefficient, with $R > 0.97$ as the criterion for an acceptable fit. All lanes from a gel were used unless a visual inspection revealed a data point to be obviously flawed relative to neighboring points. The data were normalized using the following equation:

$$\theta_{\text{norm}} = \frac{\theta_{\text{app}} - \theta_{\min}}{\theta_{\max} - \theta_{\min}} \quad (3)$$

10 2. Regulatory Moieties

A. Overview of Transcription

The major players in the regulation of gene expression within the nucleus are: the genes and their regulatory sequences that are complexed with structural proteins (e.g., histones) in chromatin; chromatin remodeling activities that allow access to a gene and its regulatory regions; regulatory proteins that instruct the transcription machinery to express (or, as in the case of repressors, prevent the expression of) the relevant genes; and the RNA-synthesizing machinery that decodes the genes. A host of other activities play a role in this process, for instance, those that facilitate elongation of paused transcripts, or those that lead to the processing of nascent transcripts and those that play a role in release of full-length transcripts.

The primary players and the events that lead to regulation of gene expression are described below. Other components involved in gene expression, such as mRNA elongation, processing, termination, or nuclear export, can also be targeted by designing synthetic regulators based on the synthetic regulatory compound motif presented herein.

Activators: Positive regulation (stimulation) of gene expression requires factors called transcriptional activators. An economical 'recruitment' model posits that activator proteins bind to DNA and recruit the transcriptional machinery to the promoter of the gene, thereby stimulating gene expression. Most activators comprise three functional modules. Of these, specificity in targeting genes is achieved by the DNA recognition module which binds to cognate DNA sequences

near a promoter of a gene and in most cases DNA binding specificity is further enhanced by dimerization. A key functional module, the activating region, is thought to bind one or more components of the transcriptional machinery. While not wishing to be bound to a particular theory, it is believed that weak interactions
5 between an activating region and several components of the machinery result in high avidity 'multi-dentate' binding. In addition, the typical activating region (e.g., those used here) are also believed to contact and recruit nucleosome modifying activities to promoters.

Repressors: These proteins appear to function to inhibit gene expression at
10 several levels. Some repressors function in part by blocking the activity of activators directly, for example, by binding to an activation domain on an activating protein in order to prevent its interaction with a component of the transcriptional machinery. Another example includes MDM-2, which not only binds to the activating region of p53, but also indirectly attenuates transcriptional activity by
15 stimulating p53's degradation via a proteolytic pathway. More recently it has been proposed that repressors are recruited to promoters where they serve to inhibit the ability of transcriptional machinery to utilize the proximal promoter by either directly interacting with the machinery and inactivating it, or indirectly by mediating changes in chromatin structure so as to prevent the components of a transcriptional
20 apparatus from interacting with DNA.

Transcriptional Machinery: The general components of the eukaryotic transcription apparatus have been described. Orphanides et al. (1996) *Genes Dev.* 10:2657-2683; and Conaway et al. (1993) *Annu. Rev. Biochem.* 62:161-190. Briefly, the transcriptional machinery for mRNA comprises the catalytic core RNA
25 polymerase II (12 subunits), several general transcription factors (TF_{II} -A, B, D, E, F, H), mediator complex (~20 Srb and Med subunits), elongator complex, co-activator proteins and several additional polypeptides, some of which remain to be defined. Most of these proteins are conserved through evolution and occur in species from yeast to humans.

30 Many of the components of the transcription machinery exist in large multi-subunit complexes which associate with the RNA polymerase II, and are known as the RNA polymerase II holoenzyme. The RNA-polymerase II holoenzyme can be

SUBSTITUTE SHEET (RULE 26)

broadly described as containing two functional parts. One part is the "catalytic core" that is required for synthesizing mRNA while the other is the mediator (Bjorkland et al. (1996) Trends Biochem. Sci. 21:335-337), a complex of approximately twenty proteins that is required for the holoenzyme to respond to
5 activators. It is believed that the holoenzyme, along with additional factors that do not associate tightly (such as TBP/TFIID and a class of proteins known as co-activators (Thompson et al. (1993) Cell 73:1361-1375; and Koleske et al. (1994) Nature 368:466-469), constitute the minimal transcriptional machinery recruited by
10 activators to most promoters in vivo. Conversely, as described above, repressors function to inhibit holoenzyme activity, and in some instances they recruit co-repressor proteins.

TFIID (Burley et al. (1996) Ann. Rev. Biochem. 65:769-799), an essential component of the transcriptional machinery, is not typically found associated with the holoenzyme, and is a target of activators and some repressors as well. It is a
15 protein complex containing about thirteen components, including TBP and TBP-associated factors (TAFs). Kim et al. (1993) Nature 365:520-527; Kim et al. (1993) Nature 365:512-520; and Dynlacht et al. (1991) Cell 66:563-576. TBP is a sequence-specific DNA-binding protein that recognizes and binds via the minor groove to a sequence known as the TATA box (consensus: 5'-TATAAAA-3') that
20 exists in the promoters of many genes. Hoopes et al. (1992) J. Biol. Chem. 267:11539-1154; and Coleman et al. (1995) J. Biol. Chem. 270:13850-13859. TFIID associates with TFIIA, which is comprised of three polypeptides. TFIIA helps TFIID bind to DNA perhaps by competing with repressors as well as displacing inhibitory domains within TAFs away from TBP. Geiger et al. (1996)
25 Science 272:830-836; and Thompson et al. (1993) Cell 73:1361-1375. TFIIB, a holoenzyme component, also interacts with the promoter DNA and binds to TBP (Nikolov et al. (1995) Nature 377:119-128; and Burley (1996) Nature 381:112-113) and it is proposed to hold the entire complex together as a single unit.

Chromatin Remodeling Machinery: In order for a gene sequestered in
30 chromatin to become available for transcription, the chromatin structure must be remodeled. Felsenfeld (1992) Nature 355:219-224; Kingston et al. (1996) Genes Dev. 10:905-92; and Kadonaga (1998) Cell 92:307-313. Chromatin remodeling

occurs through activator-mediated recruitment of at least two types of chromatin remodeling complexes. One of which comprise the histone acetyl transferases that contain proteins that acetylate certain lysine residues in the amino-terminal tails of histone proteins (Brownell et al. (1996) Curr. Opin. Genet. Dev. 6:176-184), thereby rendering DNA in a nucleosome more accessible to DNA-binding transcription factors. The second type of chromatin remodeling complex, Swi/Snf uses energy derived from ATP hydrolysis to facilitate binding of the transcriptional machinery to a particular promoter. Burns et al. (1997) Mol. Cell. Biol. 17:4811-4819; Quinn et al. (1996) Nature 379:844-847; Kwon et al. (1994) Nature 370:477-481; and Cote et al. (1994) Science 265:65-68. Activators can recruit chromatin remodeling complexes through direct binding. The viral activator VP16 has been shown to bind to components of both the multi-protein histone acetyl transferase (HAT) complex (Berger et al. (1992) Cell 70:251-265; and Candau et al. (1997) EMBO J. 16:555-565), as well as the Swi/Snf complex. In fact, TFIID, another target of VP16, was observed to display a weak HAT activity. Mizzen et al. (1996) Cell 87:1261-1270; and Wilson et al. (1996) Cell 84:235-244.

As a corollary it has been shown that certain gene-specific transcriptional repressors mediate their repressive function by recruiting histone deacetylase complexes to a target promoter. Brehm et al. (1998) Nature 391:597-601; and Magnaghi-Jaulin et al. (1998) Nature 391:601-605. Other repressors are suggested to directly bind histones and/or other similar proteins and these interactions lead to compact chromatin structures that occlude the transcriptional machinery.

The Regulatory Process: Based on current understanding, upon receipt of a signal, an activator bound to a promoter or enhancer recruits the chromatin remodeling machinery to the adjacent promoter. It then recruits the transcriptional machinery to form a pre-initiation complex at the promoter. It appears that assembly of a pre-initiation complex can require two synchronized steps: TFIID/TBP -TATA binding in concert with the association of the holoenzyme with the complex at the promoter. Stargell et al. (1996) Trends Genet. 12:311-315. For mRNA synthesis to be initiated at a particular gene, the complex must open (melt) the double helix to expose the template strands. Once mRNA initiation occurs and after a certain length of transcript is synthesized the polymerase must move away

from the promoter to continue mRNA synthesis. Certain activators such as HSF and Tat function to stimulate this stage of transcription process, possibly by recruiting the pTEFB complex which contains a kinase (Cdk9) capable of phosphorylating the largest of the 12 subunits of the polymerase. It has been reported that promoter
5 escape appears to involve hyperphosphorylation of the carboxy-terminal domain of the largest subunit of the RNA polymerase II. This hyperphosphorylation achieves two goals: first, it can provide the signal to detach the mediator complex from the catalytic core; and second, it can permit the association of RNA processing and elongator complexes with the rapidly elongating polymerase.

10 The release of the mediator and TFIID during promoter escape by the polymerase would provide a mechanistic basis for a re-initiation event by another polymerase catalytic core. Svejstrup et al. (1997) Proc. Natl. Acad. Sci. USA 94:6075-6078; and Zawel et al. (1995) Genes Dev. 9:1479-1490. It has been found that mediator complexes are limiting, whereas the catalytic machinery is more
15 abundant. Moreover, activators directly interact with both the mediator as well as TBP/TFIID, thus, they can play a major role in helping to retain the mediator and/or TFIID at the promoter. Therefore, the next transcription complex can be reassembled rapidly by only recruiting the core fragment of the RNA polymerase II holoenzyme. It is postulated that re-initiation is much more likely than initiation
20 alone to contribute significantly to rapid stimulation of gene expression. And activators must clearly play a role in (Ho et al. (1996) Nature 382:822-826) facilitating multiple rounds of transcription re-initiation.

Repression, on the other hand, requires the opposite series of events. A repressor can first directly engage an activator and mask its activating surface
25 thereby preventing its interactions with the transcriptional and chromatin remodeling machinery. As in the case of MDM-2, after masking the activating region the repressor can also directly interrupt the low-level activator-independent assembly of the transcriptional machinery at the exposed promoter. In the next set of events the repressor, such as Retinoblastoma gene product (Rb), can directly recruit histone
30 deacetylases, which then strip the acetyl groups off the lysine residues on histone tails. It is now believed that deacetylated histone H3 tails are then methylated by methyl transferases, which are also recruited by repressors. The methylated histone

SUBSTITUTE SHEET (RULE 26)

tails bind to chromatin compacting proteins such as HP-1. Thus, in a sequential manner the gene is silenced. Additional components that participate in stimulation as well as repression of a gene will no doubt be discovered in the future, and they shall also be amenable to manipulation by compounds within the scope of this invention.

A regulatory moiety is any molecule that can positively or negatively effect transcription of a target gene, other than by direct electrostatic interaction with double-stranded DNA. Representative embodiments of regulatory moieties include peptides, polypeptides, lipids, carbohydrates, and any combination thereof. A regulator moiety can be naturally occurring or be derived from or an analog of a naturally occurring molecule. Alternatively, it can be synthetic. Preferred regulatory moieties are peptides and small organic molecules.

B. Activators

An activator is any molecule that can activate transcription of a target gene. According to current understanding, an activator binds to its cognate sites in the genome and recruits the transcriptional machinery to nearby promoters; initiation of transcription then follows. Ptashne et al. (1997) Nature 386:569-577. Activators can be small molecules or peptides. For example, many protein activators are known. The peptides within such proteins that provide activation activity can be identified. Peptidomimetics of such peptides can also be generated, as can other analogs and derivatives. The key is whether the peptide or other molecule retains the desired activation function. Small molecules can be developed from rational design approaches modeled after known activators. Alternatively, or in addition, they can be identified by screening combinatorial libraries, natural product libraries, and/or libraries of already synthesized small organic molecules. Such molecules can be tested for activator function in a suitable test system. For example, one can substitute the regulatory moiety of synthetic regulatory compound already known to provide regulatory activity with a molecule of unknown activity. Such candidate molecules can be screened in an in vitro or cell-based reporter system comprising a reporter construct that carries a reporter gene the expression of which is under the control of a regulatable promoter. Candidate compounds that are found to activate transcription of the reporter gene can be retained for further study. Also, the

SUBSTITUTE SHEET (RULE 26)

particular regulatory moiety can be identified as an activator. The activator can also then be tested for activity in other compounds that employ different linkers and/or nucleic acid binding moieties, and even different regulatory moieties, or multiples of the same regulatory moiety. In this way, it will be possible to develop a library of
5 activators, the members of which can be used for various, and even multiple, applications.

C. Repressors

A repressor is any molecule that can inhibit or prevent transcription of a target gene. Repressors can be small molecules or peptides. For example, many
10 protein repressors are known. The peptides within such proteins that provide repressor activity can be identified. Peptidomimetics of such peptides can also be generated, as can other analogs and derivatives. The key is whether the peptide or other molecule retains the desired repressor function. Repressors can be obtained in ways analogous to those used to identify activators.

15 3 Linkers

For purposes of this invention, a linker is any molecule that can be used to link at least one nucleic acid binding moiety to at least one regulatory moiety in a manner that allows the nucleic acid binding moiety(ies) to retain its(their) intended nucleic acid binding function(s) and the regulatory moiety(ies) to retain its(their)
20 ability to influence transcription of the target gene. Thus, the linker should provide adequate spacing between and/or orientation with respect to the nucleic acid binding moiety and the regulatory moiety so as to allow each to retain its respective function. A linker is a polymer of backbone units e.g. methylene, propyl, and ether groups. Other backbone units include, but are not limited to, five or six membered rings.
25 One or more backbone units can include a heteroatom including, but not limited to, sulfur, nitrogen or oxygen.

Linkers used in the practice of this invention can be branched, although straight chain molecules are preferred. They can be amphipathic or aliphatic, with molecules in the latter class being preferred. Representative examples of suitable
30 linkers include polyethylene glycol, alkyl chains, and peptides.

A linker can contain one or more first reactive moieties for conjugation by a suitable chemistry to a nucleic acid binding moiety. Similarly, a linker can contain

one or more second reactive moieties for conjugation by a suitable chemistry to a regulatory moiety. The chemistry used to conjugate the first reactive moiety(ies) to the nucleic acid binding moiety(ies) can be the same or different as the chemistry used to conjugate the second reactive moiety(ies) of the linker to the regulatory moiety(ies). In certain preferred embodiments, the linker is a dendrimer with respect to second reactive moieties, in that it contains two or more such moieties and thus can be useful in linking two or more regulatory moieties having the same or different regulatory activities to a nucleic acid binding moiety.

Linkers can be conjugated to a nucleic acid binding moiety at any suitable location. Such locations include at the one or both ends of the molecule, or at an internal position. Preferably, when conjugated to the nucleic acid binding moiety, the linker is oriented such that upon interaction with the nucleic acid molecule (as mediated by the nucleic acid binding moiety) it is projected away from the nucleic acid molecule so as to avoid steric hindrance or other interaction with the nucleic acid. When the nucleic acid binding moiety is a polyamide, preferred locations for conjugating a linker thereto include the carboxy terminus, the amino terminus, and an internal amino acid residue (e.g., a β -alanine residue, a substituted pyrrole residue, an unsubstituted pyrrole residue, a substituted imidazole residue, and an unsubstituted imidazole residue). With respect to polyamides capable of assuming a hairpin conformation in the minor groove of dsDNA, certain preferred embodiments involve conjugation of the linker to the internal amino acid residue that mediates hairpin formation, for example, γ -aminobutyric acid.

Linkers useful in the practice of this invention can be cleavable, for example, by an enzyme or chemical action (e.g., photooxidation). In this way, for example, the activity of a synthetic regulatory compound can be controlled by endogenous degradative processes.

Suitable linkers can be identified from among candidate linkers by any of a number of suitable approaches. An example of one such in vitro system is as follows: a first reactive moiety on the candidate linker is used to attach the linker to a nucleic acid binding moiety known to specifically bind to a target nucleotide sequence in dsDNA in proximity to a regulatable promoter from which high level

SUBSTITUTE SHEET (RULE 26)

transcription of a reporter gene can be initiated. In this example, the first reactive moiety is used to attach the linker to a portion of the nucleic acid binding moiety in a manner that is not anticipated to substantially disrupt the nucleic acid binding moiety's ability to bind to its target nucleotide sequence. Whether such disruption occurs, or the extent of any such disruption, can be independently assessed by
5 determining the binding affinity of the nucleic acid binding moiety for its target nucleotide sequence before and after linker attachment. After obtaining a nucleic acid binding moiety-linker intermediate that retains some (at least about 10%, preferably at least 25%, and more preferably at least about 50%), substantially all (at
10 least about 70%, preferably at least about 85%, and more preferably at least about 90-95%), or all (more than 95%, preferably more than 98%) of the nucleic acid binding moiety's binding affinity and specificity for its target nucleotide sequence, the intermediate can be linked to a regulatory moiety via a second reactive moiety contained in the linker. Again, it is desirable that the linker-regulatory element
15 linkage not disrupt the biological activity of the moiety to which it is attached via the second reactive group, namely the activator or repressor. Whether such disruption also occurs, or the extent of any such disruption, can be independently assessed by determining the biological activity of a regulatory moiety before and after linker attachment. After obtaining a regulatory moiety-linker intermediate that retains
20 some (at least about 10%, preferably at least 25%, and more preferably at least about 50%), substantially all (at least about 70%, preferably at least about 85%, and more preferably at least about 90-95%), or all (more than 95%, preferably more than 98%) of the regulatory moiety's activity, that linkage mechanism is preferably selected for use generating a synthetic regulatory compound according to the invention. The
25 order of moiety conjugation to the linker can vary, and can depend on the particular nucleic acid binding moiety(ies) and/or regulatory moiety(ies) being employed.

4. Screens

The synthetic regulatory compounds of the present invention comprise at least three elements: a non-natural nucleic acid binding moiety; a linker; and a
30 regulatory moiety. As most applications the compounds of the invention concern use in cells, it is desirable to assay compounds to ensure that they are able to enter into cells and exert the desired effect. As the foregoing suggests, a synthetic

regulatory compound according to the invention that is used in an in vitro system need not exhibit cell permeability. However, in the context of cells, in order to be a synthetic regulatory compound within the scope of the invention, such compound must enter cells, and then exert the intended desired regulatory effect.

5 In general, each of the moieties involved in constructing synthetic regulatory compounds of the invention are preferably independently tested, or screened, for their ability to provide the desired function (targeted nucleic acid binding or regulatory function in the case of nucleic acid binding and regulatory moieties, respectively) or structure (in the case of linkers). For example, a nucleic acid
10 binding moiety will generally be tested to determine if it binds the desired target sequence with the requisite affinity and specificity prior to being incorporated into a conjugate including a linker and/or a regulatory moiety. Similarly, a regulatory moiety is preferably determined to have the desired regulatory effect prior to its inclusion in a conjugate comprising a nucleic acid binding moiety and a linker.

15 Screens to identify such moieties can be conducted as follows. After synthesis, whether by solid state, solution phase, or recombinant techniques, as the case may be, the particular moiety is typically tested in an in vitro format. For example, in the context of nucleic acid moieties, the nucleic acid binding moiety is exposed to nucleic acid molecules which include the intended target sequence,
20 preferably under conditions that at least approximate those expected to be encountered when the molecule is put to its intended use. It is then ascertained by any of a number of conventional methods whether the desired binding events take place, and if so, at what affinities, etc.

 In order to synthesize a moiety for testing in the first instance, any suitable
25 method can be employed. Such methods include the synthesis of a single compound by traditional methods, up through a massively parallel combinatorial approach. For example, a number of combinatorial synthetic methods are known in the art. For example, Thompson & Elman ((1996) Chem. Rev. 96:555) recognized at least five different general approaches for preparing combinatorial libraries on solid supports.
30 These were: (1) synthesis of discrete compounds; (2) split synthesis (split and pool); (3) soluble library deconvolution; (4) structural determination by analytical methods; and (5) encoding strategies in which the chemical compositions of active candidate

are determined by unique labels, after testing positive for biological activity. Synthesis in libraries in solution includes at least spatially separate synthesis, and synthesis pools. Additional descriptions of combinatorial methods are known in the art. See, e.g., Lam et al. (1997) Chem. Rev. 97:4111.

5 These approaches can be readily adapted to prepare moieties for use in accordance with the present invention, including suitable protection schemes, as necessary. After synthesis and testing of the various moieties required to make a particular compound according to the invention, they can be assembled into a synthetic regulatory molecule. After a putative synthetic regulatory compound
10 according to the invention has been generated, it can be tested in any number of in vitro or cell-based assays that elicit detectable signals in proportion with the activity of the compound. Preferably, such assays are conducted in a high throughput format. For example, use of a plurality of microtiter plates allows the simultaneous testing of large numbers of candidate compounds, be they in vitro or cell-based
15 assays. High throughput formats are often partially or fully automated, and allow 100, 1,000, 10,000, or more candidate compounds to be screened at one time. As the synthetic regulatory compounds of the invention regulate transcription, such assays typically will employ systems in which one or more reporter genes can be expressed under the control of a promoter that can be influenced by the compound.
20 Examples of suitable reporter constructs include those that encode genes such as luciferase, green fluorescent protein, CAT, etc., although any gene, the expression of which can be readily detected, can be employed.

5. Cell Permeability

25 In order to be useful in cellular contexts, synthetic regulatory compounds of the invention are preferably cell permeable, i.e., they can traverse a cell's plasma membrane and thus be internalized. Preferably, the compounds inherently have this ability. Alternatively, or in addition, they can be formulated in a composition that facilitates cell entry, for example, a liposome. When so formulated, such compositions can further comprise a cell targeting element to direct the composition
30 to a particular cell type, for example, a cell expressing a specific antigen (e.g., a disease-associated antigen) not expressed on the surface of other cells.

Cell permeability enables a compound according to the invention to enter into a cell's cytoplasm, from which it then moves into the nucleus to exert its intended regulatory effect. As is known in the art, plasma membranes present a barrier to many molecules that, if they could enter a cell, could exert a useful effect.

5 Preferably, a compound that effects an intracellular function or activity, for example, transcription of one or more particular genes, is soluble in both the aqueous compartments of the cells and organism, and preferably also in the lipid bilayers through which it must pass in order to enter cells or organelles, for example, mitochondria and chloroplasts. Nuclei, on the other hand, have large pores that
10 generally do not prevent ingress of a synthetic regulatory compound according to the invention.

Cell permeability can be assessed in a number of ways. For example, the intracellular concentration of a compound can be determined, and compared to the extracellular concentration of a compound. This can also be done over time, so that
15 the rate at which a compound is taken up can be determined.

Compositions

To be used for an agricultural or medicinal application, a synthetic regulatory compound of the invention will typically be formulated into a suitable composition. Such compositions include liquid and solid, or dry, compositions. The particular
20 composition selected will depend on a variety of factors, including the particular synthetic regulatory compound(s) to be formulated, the intended use (e.g., for an agricultural purpose (for instance, as a pesticide or as a molecular switch, such as to induce flowering) or a medical application (e.g., to treat or prevent a disease)), the method of delivery (for example, in an agricultural context, spraying or
25 broadcasting, and in a medical context, injection or oral delivery), regulatory requirements, etc.

Formulations suitable for human or non-human animal uses are among the preferred embodiments of this aspect of the invention. With regard to liquid formulations, one or more synthetic regulatory compounds preferably are suspended
30 in an aqueous carrier, for example, in an isotonic buffer solution at a suitable pH. These compositions can be sterilized by conventional sterilization techniques, or can be sterile filtered. For human or non-human animal use, the compositions can

contain pharmaceutically or veterinarily, as the case can be, acceptable auxiliary substances as required to approximate physiological conditions, such as pH buffering agents. Useful buffers include for example, sodium acetate/acetic acid buffers. The desired isotonicity can be accomplished using sodium chloride or other pharmaceutically acceptable agents such as dextrose, boric acid, sodium tartrate, propylene glycol, polyols (such as mannitol and sorbitol), or other inorganic or organic solutes. Sodium chloride is preferred particularly for buffers containing sodium ions. Many pharmaceutically acceptable carriers and their formulation are described in standard formulation treatises, e.g., Remington's Pharmaceutical Sciences by E.W. Martin. See also, Wang, et al. (1988) J. Parenteral Sci. Tech., Technical Report No. 10, Supp. 42:2S.

The synthetic regulatory compounds of the invention can also be formulated as pharmaceutically acceptable salts and/or complexes thereof. Pharmaceutically acceptable salts are non-toxic salts at the concentration at which they are administered. The preparation of such salts, including addition salts, can facilitate pharmacological or other use by altering the physical-chemical characteristics of the active ingredient without preventing the synthetic regulatory compound from exerting its intended physiological effect. Pharmaceutically acceptable salts of compounds of this invention include salts of their conjugate acids or bases. Exemplary suitable counterions for conjugate acid salts include the chlorides, bromides, phosphates, sulfates, maleates, malonates, salicylates, fumarates, ascorbates, benzenesulfonates, methanesulfonates, p-toluenesulfonates, cyclohexylsulfonates, lactates, malates, citrates, acetates, tartrates, succinates, glutamates, sulfamates, quinates, and the like, in particularly those salts which are FDA acceptable. A conjugate acid salt can be formed by contacting a synthetic regulatory compound in the free acid or base form with a sufficient amount of the desired base or acid, for example, hydrochloric acid, sulfuric acid, phosphoric acid, sulfamic acid, acetic acid, citric acid, lactic acid, tartaric acid, malonic acid, methanesulfonic acid, ethane sulfonic acid, benzene sulfonic acid, p-toluenesulfonic acid, cyclohexyl sulfamic acid, and quinic acid. Such contacting can typically involve reacting the free acid or base forms of a synthetic regulatory compound with one or more equivalents of the appropriate base or acid in a solvent or medium in

which the salt is insoluble, or in a solvent such as water that is then removed, for example, in vacuo, by freeze-drying, or by ion exchange. A conjugate acid or base form of a compound of this invention is considered equivalent to the free base form (or the free acid form, as the case may be) for the purposes of the claims of this invention.

Carriers and/or excipients can also be included in compositions of the invention. Representative examples of carriers and excipients include calcium carbonate, calcium phosphate, various sugars such as lactose, or types of starch, cellulose derivatives, gelatin, vegetable oils, polyethylene glycols, and physiologically compatible solvents. Excipients such as polyhydric alcohols and carbohydrates share the same feature in their backbones, i.e., -CHOH-CHOH- . Useful polyhydric alcohols include such straight-chain molecules as sorbitol, mannitol, inositol, glycerol, xylitol, and polypropylene/ethylene glycol copolymer, as well as various polyethylene glycols (PEG) of various molecular weights, including molecular weights of 200, 400, 1450, 3350, 4000, 6000, and 8000. Carbohydrates, for example, mannose, ribose, trehalose, maltose, glycerol, inositol, glucose, galactose, arabinose, can also be included.

If desired, liquid compositions can be thickened with a thickening agent such as methylcellulose. They can be prepared in emulsified form, either water in oil or oil in water. Any of a wide variety of suitable emulsifying agents, including pharmaceutically acceptable emulsifying agents, can be employed including, for example, acacia powder, a non-ionic surfactant (such as a Tween), or an ionic surfactant (such as alkali polyether alcohol sulfates or sulfonates, e.g., a Triton).

In general, compositions of the invention are prepared by mixing the ingredients following generally accepted procedures. For example, the selected components can be simply mixed in a blender or other standard device to produce a concentrated mixture that can then be adjusted to the final concentration and viscosity by the addition of water or thickening agent and possibly a buffer to control pH or an additional solute to control tonicity.

Compositions of the invention will typically be provided in a dosage form or formulation. Any suitable dosage form can be employed, and different dosage forms can be used for different applications. Exemplary formulations within the

scope of the invention include a parenteral liquid dosage form, a lyophilized or freeze-dried unit-dosage form, controlled or sustained release formulations, wherein an effective amount of the active ingredient is released over time, and modifications of these dosage forms that are useful in practicing certain aspects of the invention.

- 5 Such dosage forms can be administered to a patient via a variety of routes, including oral, nasal, buccal, sublingual, intratracheal, ocular, transdermal, and pulmonary delivery.

Formulations that support a parenteral liquid dosage form (for intravenous, intramuscular, interperitoneal, peripheral, or subcutaneous injection or infusion, for
10 example) are those in which the active ingredient(s) are stable, and typically the solvent has adequate buffering capacity to maintain the pH of the solution over the intended shelf life of the product, and can optionally include a preservative. The dosage form should be either an isotonic and/or an iso-osmolar solution to either facilitate stability of the active ingredient or lessen the pain on injection or both.

- 15 Oral delivery can be accomplished in a variety of ways, for example, by liquid (including gel caps) or solid dosage forms, and certain preferred embodiments concern pharmaceutical formulations intended for oral administration. Such formulations can be prepared as solid dosage forms. Particularly preferred solid dosage forms are pills, e.g., capsules, tablets, caplets, or the like suitable for oral
20 ingestion. Solid dosage forms typically contain inert ingredients (e.g., carriers and excipients, as described above) along with the active ingredient to facilitate tablet formation. Numerous capsule manufacturing, filling, and sealing systems are well-known in the art, and can be used to make pills of any desired size. Preferred capsule dosage forms can be prepared from gelatin or starch. After making a
25 capsule dosage form, if desired, it can be coated with one or more suitable materials. For example, one or more enteric coatings can be applied to prevent gastric irritation, nausea, or to prevent the active ingredient from being destroyed by acid or gastric enzymes, or to target a particular gastrointestinal region.

- Formulations that support pulmonary and/or intra-tracheal dosage forms
30 include preserved or unpreserved liquid formulations and/or dry powder formulations, as can be used in any suitable device for delivering the composition to the lung, e.g., a metered dose inhaler, nebulizer, or dry powder inhaler. Dissolvable

gels and/or patches are useful to facilitate buccal delivery, and can be prepared from various types of starch and/or cellulose derivatives. Sublingual delivery can be supported by liquid formulations or by solid dosage forms that dissolve under the tongue.

5 After a dosage form is prepared, it is typically packaged in a suitable material. For pill or tablet dosage forms, the dosage forms can be packaged individually or bottled en masse.

 The effective dosage of a synthetic regulatory compound of the invention will vary depending upon a variety of factors, including the compound itself, its
10 intended application, etc. When the compound is a pharmaceutical or veterinary drug, the particular dose and dosing regimen can be determined by the attending clinician, and can be further dependent upon such factors as the age, weight, and condition of the patient.

 Those skilled in the art will be able to use the preceding information to
15 prepare appropriate formulations for delivery of compositions comprising a synthetic regulatory compound of the invention. Other necessary information is known in the art and can be utilized to prepare appropriate formulations.

Applications

 Misregulation of transcriptional cascades is often the cause of disease; thus,
20 gene-specific, particularly regulatory element-specific, modulation of gene expression mediated by, for example, cell permeable synthetic regulatory compounds offer the opportunity to treat or prevent disease that arise due to ectopic gene expression. An important application of the compounds of the invention is in the study of gene expression at a mechanistic level in cell-free systems and in living
25 cells. Nucleic acid binding moieties (e.g., polyamides) employed in the compounds of the invention provide a level of promoter- or other regulatory element selectivity that can not be afforded by natural or chimeric transcriptional factors.

 In certain preferred embodiments of the invention, compositions comprising a synthetic regulatory compound can be used to modulate physiological processes in
30 vivo. In animals, particularly humans, primates, and domestic animals, one can affect development by controlling the expression of particular genes, modify physiological processes, such as accumulation of fat, growth, response to stimuli,

etc., and treat or prevent disease. Domestic animals include bovine, canine, equine, feline, murine, ovine, and porcine animals, as well as fish and birds. The compositions of the invention can also be used agriculturally, for example, to control plant and animal pests, to affect gene regulation in plants, particularly commercially important crops.

The synthetic regulatory compounds of the invention can be used for the treatment or prevention of various disease states. The subject compositions can be used, for example, therapeutically, to inhibit or activate the expression of one or more genes, which can change the phenotype of cells, either endogenous or exogenous to the host or patient, where the phenotype is detrimental. For example, synthetic regulatory compounds can be developed that contain a nucleic acid binding moiety that specifically binds to nucleic acids (e.g., double-stranded genomic DNA) in pathogens (i.e., viruses and pathogens of either eukaryotic or prokaryotic origin) that are involved in the regulation of expression of certain pathogenic genes, for example, genes required for replication or virulence. Alternatively, the nucleic acid binding moiety can be chosen to target a synthetic regulatory compound to a unique sequence of the pathogen that is not found in the genome of the pathogen's host. Thus, by inhibiting the expression of housekeeping or other genes of bacteria or other pathogens, particularly genes specific to the pathogen, one can provide for inhibition of proliferation and/or virulence of the particular pathogen.

Similarly, where a gene may be essential to proliferation or protect a cell from apoptosis, where such cell exhibits undesired proliferation, the subject compositions can be used to inhibit expression of the gene by inhibiting transcription or chromatin remodeling thereof. Alternatively, where a disease phenotype is caused by under-expression of a gene, its expression can be activated by providing a compound according to the invention. An important application of the invention's synthetic regulatory compounds is in the prevention and treatment of cancer. For example, expression of specific oncogenes can be inhibited or prevented. Similarly, expression of genes inappropriately up-regulated in cancer cells can also be inhibited or prevented. Also, other genes whose expression is essential to the maintenance of an immortal phenotype, e.g., the genes coding for the RNA and protein components of telomerase, can be down-regulated using a

synthetic regulatory compound. An alternative strategy involves the activation of genes that give rise to an apoptotic cascade or whose expression is down-regulated as part of the tumorigenic process. These approaches to regulating transcription will find application in situations such as cancers, such as sarcomas, carcinomas and

5 leukemias, restenosis, psoriasis, lymphopoiesis, atherosclerosis, pulmonary fibrosis, primary pulmonary hypertension, neurofibromatosis, acoustic neuroma, tuberous sclerosis, keloid, fibrocystic breast, polycystic ovary and kidney, scleroderma, inflammatory diseases such as rheumatoid arthritis, ankylosing spondylitis, myelodysplasia, cirrhosis, esophageal stricture, sclerosing cholangitis,

10 retroperitoneal fibrosis, etc. Inhibition or activation, as the case may be, can be associated with one or more specific growth factors, such as the families of platelet-derived growth factors, epidermal growth factors, transforming growth factor, nerve growth factor, fibroblast growth factors, e.g., basic and acidic, keratinocyte fibroblast growth factor, tumor necrosis factors, interleukins, particularly interleukin

15 1, interferons, etc. In other situations, one can wish to inhibit a specific gene that is associated with a disease state, such as mutant receptors associated with cancer, or inhibit the arachidonic cascade, expression of various oncogenes, including transcription factors, such as ras, myb, myc, sis, src, yes, fps/fes, erbA, erbB, ski, jun, crk, sea, rel, fms, abl, met, trk, mos, Rb-1, etc. Other conditions of interest for

20 treatment with the subject compositions include inflammatory responses, skin graft rejection, allergic response, psychosis, sleep regulation, immune response, mucosal ulceration, withdrawal symptoms associated with termination of substance use, pathogenesis of liver injury, cardiovascular processes, neuronal processes, and, in particular, where specific T-cell receptors are associated with autoimmune diseases,

25 such as multiple sclerosis, diabetes, lupus erythematosus, myasthenia gravis, Hashimoto's disease, cytopenia, rheumatoid arthritis, etc., the expression of the undesired T-cell receptors can be diminished, so as to inhibit the activity of the disease-associated T-cells. In cases of reperfusion injury or other inflammatory insult, one can provide for inhibition of enzymes associated with the production of

30 various factors associated with the inflammatory state and/or septic shock, such as TNF, enzymes that produce singlet oxygen, such as peroxidases and superoxide dismutase, proteases, such as elastase, INF- γ , IL-2, factors that induce proliferation

of mast cells, eosinophils, IgG, IgE, regulatory T cells, etc., or modulate expression of adhesion molecules in leukocytes and endothelial cells.

Compounds of this invention can be screened for their in vitro activities against different species of bacteria and fungi. The minimal inhibition concentration (MIC) of these compounds was determined using the National Committee for Clinical Laboratory Standards (NCCLS) broth microdilution assay in microtiter plates, as set forth in: (1) the guidelines of the National Committee for Clinical Laboratory Standards (NCCLS) Document M7-A4 (NCCLS, 1997); (2) the guidelines of the National Committee for Clinical Laboratory Standards (NCCLS) Document M11-A4 (NCCLS, 1997); and (3) the guidelines and reference method of the National Committee for Clinical Laboratory Standards (NCCLS) Document M27-T (NCCLS, 1995). For antifungal assays, the method recommended in Murray, PR., 1995 Manual of clinical microbiology (ASM Press, Washington, DC.), was employed. A variety of Gram positive and Gram-negative bacteria (aerobes and anaerobes) as well as yeasts and filamentous fungi were tested. These organisms included *Staphylococcus* spp., *Streptococcus* spp., *Enterococcus* spp., *Corynebacterium* spp., *Listeria* spp., *Bacillus* spp., *Micrococcus* spp., *Peptostreptococcus* spp., *Clostridium* spp., *Propionibacterium* spp., *Escherichia* spp., *Pseudomonas* spp., *Haemophilus* spp., *Candida* spp., *Cryptococcus* spp., *Aspergillus* spp., *Trichophyton* spp., *Paecilomyces* spp., *Saccharomyces* spp. and *Fusarium* spp. In addition, some drug resistant microbes were also evaluated with this assay. Other pathogenic bacteria against which compounds of this invention can be effective include *Acinetobacter* spp., *Alcaligenes* spp., *Campylobacter* spp., *Citrobacter* spp., *Enterobacter* spp., *Proteus* spp., *Salmonella* spp., *Shigella* spp., *Helicobacter* spp., *Neisseria* spp., *Vibrio* spp., *Bacteroides* spp., *Prevotella* spp., *Mycoplasma* spp., *Mycobacteria* spp., and *Chlamydia* spp.

Other opportunities for use of the subject synthetic regulatory compounds include modulating the level of expression of genes coding for receptors, ligands, enzymes, changing phenotype of cells, modifying the response of cells to drugs or other stimuli, e.g., enhancing or diminishing the response, and inhibiting or activating the expression of one of two or more alleles.

To accomplish the above representative treatments and therapies, individual and multiple compounds can be employed, directed to the same dsDNA region, but different target sequences, contiguous or distal, or different DNA regions, depending upon the number of genes which one wishes to modulate the expression of. The
5 subject compositions can also be used as a sole therapeutic agent or in combination with other therapeutic agents. Depending upon the particular indication, other drugs can also be used, such as antibiotics, antisera, monoclonal antibodies, cytokines, anti-inflammatory drugs, and the like. The subject compositions can be used for acute or chronic situations, where a particular regimen is devised for the treatment of
10 the patient.

The following examples are provided to assist in understanding the present invention. The examples and experiments described below should not, of course, be construed as specifically limiting the invention and such variations of the invention, now known or later developed, which would be within the purview of one skilled in
15 the art in view of the description provided herein.

EXAMPLE 1

Preparation of Thioesters for Ligating Peptides With Non-Native Substrates

This example provides a one-pot procedure for the synthesis of thioesters from primary amines. Polyamides containing one or more primary amines were
20 prepared by solid-phase synthesis using standard methods (see U.S. Patent Nos. 6,090,947 and 5,998,140) and reacted with thiolane-2,5-dione followed by alkylation with benzyl bromide to produce the target thioesters in good yield. The thioesters thus prepared were conjugated to peptides to produce synthetic regulatory compounds according to the invention using the "native chemical ligation" method.
25 This flexible synthetic procedure provides a ready route to both natural and unnatural substrates for chemical ligation reactions.

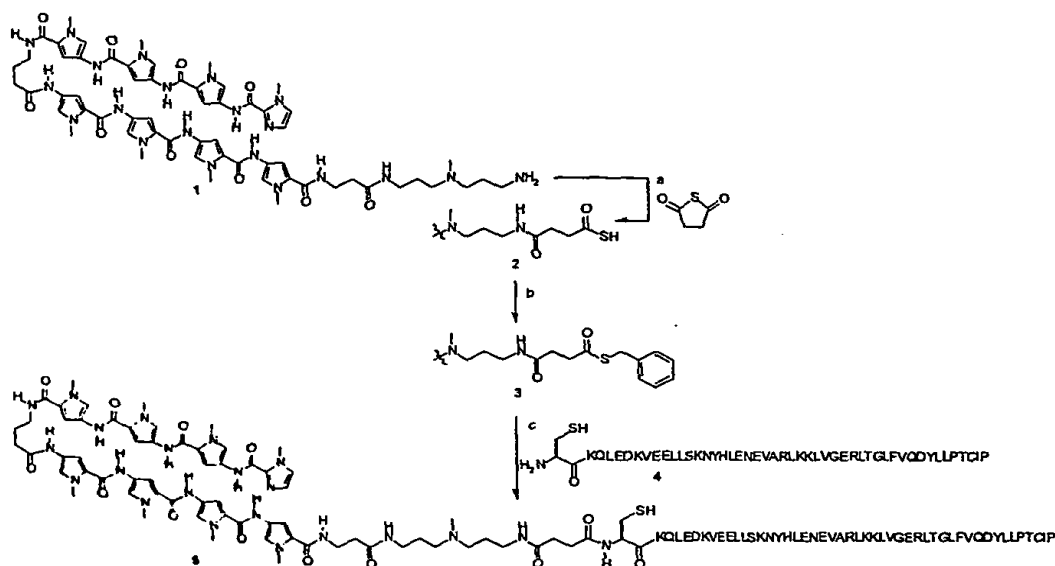
In order to prepare large numbers of putative synthetic regulatory compounds, a streamlined synthetic process was desired. Of the many coupling techniques available to prepare peptide conjugates, the most versatile was the
30 "native chemical ligation" procedure originally developed for the synthesis of proteins too large to be accessed by standard solid-phase synthesis approaches. Dawson et al. (1994) Science 266:776-779. In this reaction, a peptide containing a

carboxy-terminal thioester and a peptide with an amino-terminal cysteine are combined in denaturing buffer. Upon transesterification of the thioester with the cysteine thiol, an S→N acyl shift takes place to generate a ligated product in which the two halves are now connected by an amide bond. The product recovery of this sequence is generally good, and the facility of the reaction appears sequence independent. Total syntheses of many natural and modified proteins have been reported using this method. Cotton et al. (1999) Chem. Biol. 6:R247-R256.

To adapt this powerful reaction for present purposes, it was necessary to prepare polyamides containing the requisite thioester functional group. Available methods for the preparation of suitable thioesters range from biochemical approaches (Muir et al. (1998) Proc. Natl. Acad. Sci. USA 95:6705-6710; and Welker et al. (1999) Biochem. Biophys. Res. Commun. 254:141-151) to solid-phase synthesis using modified resins. See, e.g., Camarero et al. (2000) Lett. Peptide Sci. 7:17-21; Shin et al. (1999) J. Am. Chem. Soc. 121:11684-11689; Ingenito et al. (1999) Tetrahedron Lett. 121:11369-11374; Schwabacher et al. (1993) Tetrahedron Lett. 34:1269-1270; Clippingdale et al. (2000) J. Pept. Sci. 6:225-234; Hojo et al. (1991) Bull. Chem. Soc. Jpn. 64:111-117; Yamashiro et al. (1988) Int. J. Peptide Protein Res. 31:322-334; and Yamashiro et al. (1981) Int. J. Peptide Protein Res. 18:385-392. However, all previous methods provided products with thioesters at the carboxy-terminal position, severely restricting their utility in the context needed to generate the desired range of compounds useful in the practice of this invention, as, in certain embodiments, it was desired to prepare compounds with multiple peptides attached to one polyamide. Accordingly, it was necessary to develop a synthetic approach that would allow such flexibility.

Polyamides were prepared by solid-phase synthesis (Baird et al. (1996) J. Am. Chem. Soc. 118:6141-6146; and U.S. Patent No. 6,090,947), and a functional group that is straightforward to introduce at one or more positions is a primary amine. Therefore, an electrophile was needed that would produce a thioester or thioacid when reacted with an amine. Thiolane-2,5-dione was readily available from succinic anhydride by reaction with sodium sulfide (Kates et al. (1995) J. Heterocyclic Chem. 32:971-978), and thus was tested.

Scheme I



Conditions: (a) DIPEA, NMP, rt. (b) 100 mM NaOAc (pH 5.2), benzyl bromide, 0 °C. (52%, 2 steps).
 (c) 100 mM potassium phosphate buffer (pH 7.3), 6 M Gn-HCl, 5% NMP, 5% PhSH, 4 d, rt (29%).

Referring to Scheme I, to test this approach, hairpin polyamide **1**, containing a primary amine, was prepared by solid-phase synthesis. Trauger et al. (1996; Baird et al.; and U.S. Patent No. 6,090,947. Polyamide **1** and thiolane-2,5-dione were combined in N-methylpyrrolidinone (NMP) with N,N-diisopropylethylamine (DIPEA) at ambient temperature and reaction progress monitored by analytical reversed-phase HPLC (Scheme 1). Polyamide **1** was rapidly consumed and thioacid **2** isolated by ether precipitation. Alternatively, thioester **3** was produced in a “one-pot” conversion by lowering the pH of the initial reaction mixture with pH 5.2 NaOAc buffer and adding benzyl bromide. As monitored by analytical HPLC, the conversion from **1** to **3** was virtually quantitative. A typical reaction procedure was conducted as follows: to a solution of 18.6 μ moles (23.5 mg) polyamide **1** in 0.600 mL NMP was added thiolane-2,5-dione (25.0 μ L of a 1.00 M solution, 25.0 μ moles), followed by 9.70 μ L (55.6 μ mol) of DIPEA. After 10 min., conversion to thioacid **2** appeared complete. The reaction mixture was diluted with 900 μ L 100 mM NaOAc buffer (pH 5.2) and cooled to 4°C, necessary to prevent the formation of dialkylation products in the subsequent step. Benzyl bromide (55.8 μ mol, 6.70 μ L) was added with thorough mixing and after an additional 10 min., the thioester **3** was isolated by semi-preparatory reversed-phase HPLC as a pale yellow powder in

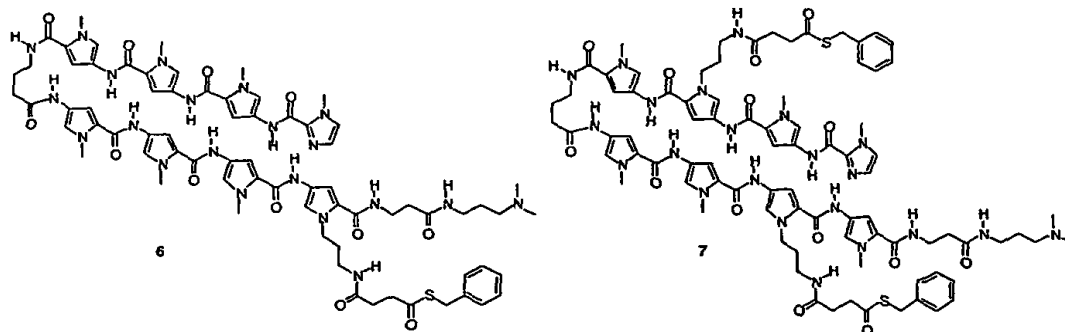
SUBSTITUTE SHEET (RULE 26)

52% overall yield (14.2 mg, 9.66 μ mol).

Data for thioester 3: ^1H NMR (500 MHz, DMSO- d_6): δ 1.58-1.68 (m, 2), 1.72-1.82 (m, 4), 2.28-2.36 (m, 4), 2.41 (t, 2, J = 6.6 Hz), 2.72 (d, 2, J = 4.9 Hz), 2.84 (t, 2, J = 6.8 Hz), 2.94-3.0 (m, 2), 3.04-3.10 (m, 4), 3.16 (tt, 2, J = 6.9, 13.9), 3.21 (m, 2), 3.80 (s, 3), 3.81 (s, 3), 3.83 (s, 3), 3.84 (s, 3), 3.84 (s, 3), 3.85 (br s, 5), 3.91 (d, 3, J = 2.0 Hz), 3.99 (s, 3), 4.10 (s, 3), 6.84 (d, 1, J = 1.7 Hz), 6.88 (d, 1, J = 2.0 Hz), 6.91 (d, 1, J = 2.0 Hz), 7.05 (br s, 3), 7.16-7.18 (m, 3), 7.21-7.30 (m, 7), 7.39 (s, 1), 7.86 (t, 1, J = 5.6 Hz), 7.99-8.10 (m, 6), 9.09 (br s, 1), 9.84 (s, 1), 9.89 (s, 1), 9.90 (s, 1), 9.91 (s, 1), 9.93 (s, 1), 9.96 (s, 1), 10.46 (s, 1). MALDI-TOF MS [M+H] (monoisotopic) calculated 1470.7, observed 1470.7.

Thioester 3 underwent reaction under standard "native chemical ligation" conditions with peptides having an amino-terminal cysteine (such as compound 4, below) to produce compound 5. Due to the hydrophobicity of thioester 3, it was necessary to dissolve the thioester in NMP (5-10% of reaction volume) prior to addition of aqueous denaturing buffer. Compound 5 was recovered in 29% yield (after reversed-phase HPLC purification) under these reaction conditions (MALDI-TOF MS [M+H] (average mass) calculated 6935.0, observed 6934.8).

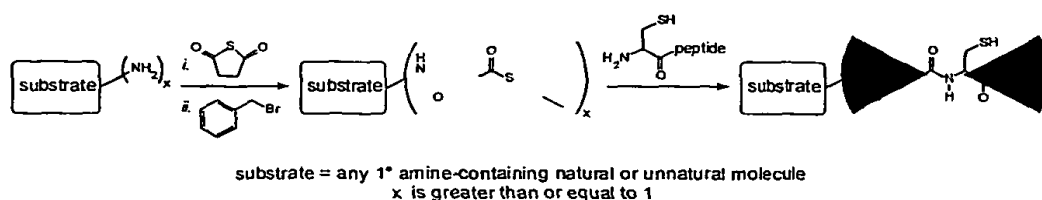
To gauge the generality of this method for preparing polyamides with multiple thioesters at diverse positions within the structure, compounds containing one or more internal primary amine(s) were used to provide the requisite



functional group handles. The above reaction sequence was applied to afford thioesters such as compounds 6 and 7. MALDI-TOF MS [M+H] (monoisotopic) 6: calculated 1470.7, observed 1470.7; 7: calculated 1719.7, observed 1719.8. Compound 6 was isolated in 55% yield after HPLC purification. Incorporation of

two thioesters was as facile as one, with conversion to product nearly quantitative as determined by analytical HPLC to furnish thioester 7 in 30% yield. In all cases, subsequent conjugation via "native chemical ligation" to peptides containing amino-terminal cysteine residues proceeded readily (5-45% yields of isolated conjugate).

- 5 In summary, the one-pot reaction sequence described herein allows rapid access to thioester-modified products suitable for "native chemical ligation" reactions (Scheme 2).



- 10 Most notably, this procedure requires only a primary amine and can be used to install multiple functional groups. This sequence works equally well for the functionalization of α -amino acids. While this example describes the use of this method to prepare polyamide-linker-regulatory compounds, it is extendable to the preparation of other chimeric target molecules. Finally, although the thioester
- 15 products function well in the native chemical ligation reaction, other ligation approaches are equally applicable such as the Staudinger ligation or alkylation of thioacids. Schnölzer et al. (1992) Science 256:221-225.

Example 2

20 Gal4 dimerization domain

- This Example describes the combination of a weakly dimerizing Gal4 domain (residues 73-100 of Gal4) as a linker to connect the polyamide and activator moieties of Example 3 (YLLPTCIP ("XL" SEQ ID NO: 2)) in order to make synthetic regulatory compounds of the invention, as well as such compounds that
- 25 instead comprise a flexible polyethylene glycol linker. XL is described in Lu et al. (2000) Proc. Natl. Acad. Sci. USA 97:1989-1992. The synthetic process described in Example 1 was used to generate each of three compounds (7-9, Figure 1 SEQ ID NOs: 1, 2 and 3) tested in the course of the experiments described below.

As shown in Figure 1A, compounds 7, 8, 9 had the same structure, except that the linker domain of each compound was different. Compound 7 did not contain a linker per se; instead, the activator peptide was linked directly the C-terminus via a Cys residue during the native ligation procedure. Conjugate 8
5 contained a PEG linker, and compound 9 contained as a linker the weak dimerization domain comprising residues 73-100 of Gal4. Each of the compounds was generated in moderate to good yield using the synthetic process of Example 2, and each was sufficiently water soluble to be purified and characterized, although to a lesser degree than the polyamide moiety alone. Each conjugate had a binding
10 affinity for the target nucleotide sequence (the same site for each compound) that was at least about 10-fold less than that of the polyamide moiety alone for its cognate target sequence. Compound 9 bound inverted target dimer site (Figure 1B) with an approximately 4-fold greater affinity ($K_a = 9 \times 10^7 M^{-1}$ vs. $2 \times 10^7 M^{-1}$), indicating that a favorable interaction occurs between adjacent compounds. The use
15 of organic solvents in DNase I footprinting experiments was necessary.

Example 3

Compounds Comprising AH Activators and Gcn4 Linkers Attached Via Internal Residues

This example describes the synthesis and testing of several synthetic
20 regulatory compounds according to the invention that also employ polyamides as the nucleic acid binding moiety. In several of these compounds, the linker is attached at an internal pyrrole unit rather than at the C-terminus of the polyamide. See Figure 2. Also, the use of an acidic activator peptide (PEFPGIELQELQELQALLQQ ("AH", SEQ ID NO: 5) is described. The AH activator increases water solubility, and also
25 contacts different components of the transcriptional machinery as compared to the small hydrophobic XL peptide. Finally, alternative dimerization domain (Gcn4 251-281) is described (compounds 10 (SEQ ID NO: 4), 11, 13 and 14, Figure 2). This dimerization domain is derived from the leucine zipper region of the yeast protein Gcn4. It is relatively small (about 30 residues) and soluble. Gcn4 conjugates
30 containing XL (conjugates 10 and 13) also included amino acid residues 90-100 of Gal4 believed to be necessary for full function. Finally, a series of compounds comprising two activator peptides linked to a single non-natural DNA binding

moiety, thus an intramolecular dimerization domain, were also synthesized (compounds 18-20, Figure 2).

Each of compounds 10-20 was prepared by joining a linker-activator peptide conjugate to a polyamide via native ligation, as described in Example 1, Scheme 1.

5 All of the compounds were prepared in low (5%) to good (50%) yield, and their respective identities confirmed by MALDI-TOF mass spectrometry.

DNase I footprint titrations were performed to determine DNA binding affinities. The DNA fragments used for the titrations were ^{32}P -labeled restriction fragments containing an inverted dimer binding site separated by 5 or 7 base pairs, as well as monomer sites. Each of the compounds bound DNA with reasonable ($K_a > 10^7 \text{M}^{-1}$) affinity. The conjugates containing the Gcn-4 dimerization domain exhibited much improved solubility characteristics and bound DNA with greater affinity than the corresponding Gal4 conjugates. Compounds containing a Gcn4(251-281) linker (compounds 10, 11, 13, and 14) bound a palindromic dimer site with about 5 fold higher affinity than the corresponding monomer site.

Each of the conjugates was separately tested in an in vitro transcription assay derived from yeast nuclear extracts to assess transcription activation function. The template DNA used in the reactions contained three palindromic binding sites 50 base pairs upstream of the transcription start site. See Figure 3. In these assays, compounds 9 and 17-20 did not activate transcription, and in some cases, inhibition or levels lower than basal were observed.

Assays to which compounds 9, 15, and 16 were added exhibited measurable (2-3 fold) levels of transcription above basal levels. Compounds 10, 11, 13, and 14 activated transcription efficiently (5-10 fold) on three different templates that contained palindromic binding sites, and the size of the palindromic binding sites appeared to affect transcription levels only slightly.

Preliminary experiments with other compounds indicated that a dimerization region could be dispensed with. Compound 11 was the most active, giving consistently high levels of activation. Minor effects of binding site size were also observed, particularly for compounds 10 and 13, which each contain XL as the activator peptide. Compound 12 was a weak activator in these assays, presumably due to an interruption of the helix proposed to form at residues 86-96 by attachment

SUBSTITUTE SHEET (RULE 26)

of the Gcn4(251-281) region which could significantly disrupt presentation of XL. Matching the phasing of the two helices can therefore improve the activity of such compounds (i.e., compounds wherein the linker and activator moieties are both peptides).

- 5 The order of addition of the various assay components has some effect upon the levels of transcription observed. When conjugates were pre-incubated with nuclear extracts prior to the addition to the template DNA, activated transcription was detected at conjugate concentrations of 5-50 nM. When conjugates were pre-incubated with template DNA for 1.5 hr. (as in the DNase I footprinting
10 *experiments*) before nuclear extract was added, activation was not observed at concentrations lower than 50 nM, and higher overall levels (7-15-fold) of activated transcription were obtained at higher concentrations (500 nM).

Example 4

Compounds Comprising Non-Dimerizing Linkers

- 15 This Example describes the design, synthesis, and testing of a series of compounds (21-28 see Figure 4) in which the Gcn4-derived linker portion (Gcn4(251-281)) of compound 11 (see Example 5) was replaced with a "scrambled" peptide in which the amino acid residues of the linker were re-ordered so as to disrupt the helix-forming propensity of the coiled-coil motif of native peptide
20 (compound 21) believed to participate in dimerization of Gcn4 molecules; the polyamide and activating domains were the same as those in compound 11. To further probe the relative importance of dimerization versus projection from DNA, compounds in which Gcn4(251-281) were replaced by one polyethylene glycol linker (compound 22), two polyethylene glycol linkers (compound 23), or no linker
25 (compound 24) were constructed. Compound 22 was predicted to orient the AH activation domain away from the double helix and project approximately one-half the distance provided by the Gcn4(251-281) linker, whereas compound 23, which comprised two polyethylene linkers, was predicted to position the AH activation domain away from the DNA at approximately the same distance as the Gcn4(251-
30 281) linker. Control compounds (compounds 25-28; see Figure 4) were also generated.

Each of the compounds was prepared by methods described in the previous examples, and their respective identities were confirmed by MALDI-TOF mass spectrometry. In vitro transcription assays were performed using a reporter construct that separated the inverted repeat of the nucleotide sequence targeted by the polyamide (the same hairpin polyamide used in the compounds described in
5 Example 3) by seven nucleotides. The inverted repeat was replicated three times, with the 3'-most repeat being spaced 50 base pairs from the TATA box of the AdML promoter, which was used to regulate expression of luciferase in a g-less expression cassette.

10 Compound 11 was observed to activate expression of the reporter gene 25-40-fold over basal levels at compound concentrations of 250-350 nM and transcription times of 30-40 minutes. Control compounds 25-28 did not produce detectable levels of expression of the reporter gene product under comparable conditions, as expected. Compound 21 activated transcription 1.5-2-fold over basal
15 levels. In contrast, when AH was attached directly to the polyamide with no intervening linker (compound 24), only minimal (2-3-fold) activation was observed. On the other hand, inclusion of a single polyethylene glycol flexible linker (compound 22) resulted in activation levels (10-fold) approximately half those generated by compound 11. Finally, on a template containing three "mismatch"
20 palindromic binding sites, the strongest activator (compound 11) activated transcription minimally (less than 2-fold), and only at high concentration (500 nM).

The compounds described in this Example demonstrate that a discrete dimerization domain is not essential for activator function. That compound 11, which comprises a leucine zipper region derived from Gcn4, exhibits stronger
25 activating potential than compounds containing a flexible linker leads to the conclusion that projection from DNA is the crucial function of the linker region. Furthermore, when the capacity to project AH from DNA via coiled-coil formation was removed (compound 21), the function of the conjugate as an activator was also severely compromised. Finally, transcription activation by these compounds
30 demonstrates that the DNA binding domain of an activator plays no fundamental role in activation other than specifying the location of DNA binding.

EXAMPLE 5

Activation of gene expression by small molecule transcription factors

As discussed elsewhere herein, naturally occurring eukaryotic transcriptional activators typically, at a minimum, comprise a dsDNA binding domain and a
5 separable activation domain, and most activator proteins also contain a dimerization module between the DNA binding and activation domains. This example describes preferred embodiments of a class of synthetic regulatory compounds according to the invention that mimic natural transcription factors, namely compounds comprising a dsDNA binding domain comprised of a polyamide, a peptide-based
10 activation domain derived from a designed or naturally occurring transcription activating protein, and either a peptidic or polyethylene glycol linker (diglycolic anhydride).

As shown in this example, such compounds can mediate high levels of DNA site-specific transcriptional activation in vitro. A representative example of such
15 molecules had a molecular weight of about 4.2 kDa, and contained a sequence-specific DNA binding polyamide to serve as the DNA binding region, a non-peptide linker instead of a dimerization peptide, and an activating region (here, a designed peptide) designated "AH", that comprised the following amino acid sequence: PEFPGIELQELQELQALLQQ (SEQ ID NO: 5, Giniger et al. supra). Because
20 synthetic polyamides can be designed to recognize any specific double-stranded nucleic acid sequence, these results demonstrate that synthetic regulatory compounds can be designed to up-regulate the expression of any specified gene.

The compounds described in this Example incorporate non-natural or synthetic counterparts for each of the functional modules typically found in naturally
25 occurring activator proteins. In the compounds, the protein-based DNA binding module was replaced with a hairpin polyamide composed of *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) amino acids that binds in the minor groove of DNA (Figure 4). The hairpin polyamide selected for the present study, ImPyPyPy- γ -PyPyPyPy- β -Dp (where γ is γ -aminobutyric acid, β is β -alanine, and Dp is
30 dimethylaminopropylamide; polyamide 1), binds the target nucleotide sequence 5'-TGTTAT-3' with a dissociation constant (K_D) of 1.1 nM. Initially, a palindromic

binding site containing this sequence as an inverted repeat separated by 7 bp was targeted. A peptidic dimerization element known to form a coiled-coil, residues 251-281 of the yeast protein Gcn4 (Ellenberger et al. (1992) Cell 71:1223-1237), was used to link the polyamide to the AH activation domain.

5 **Synthesis of Conjugates 3, 4, and 5 (Figure 5).** Polyamide 1 (Example 1) was prepared according to established protocols as described above and then was combined with 1.2 equivalent (eq.) thiolane-2,5-dione (Kates et al. (1995)) and 3 eq *i*-Pr₂EtN in 1-methyl-2-pyrrolidinone at a final concentration of 10 mM. After 15 min., 1.5 vol. of 100 mM NaOAc (pH 3.2) were added followed by 3 eq. of benzyl
10 bromide. After an additional 15 min., the reaction mixture was subjected to purification by reversed-phase HPLC, and the appropriate fractions were concentrated to isolate compound 2 (Example 1) (53%) as a white powder. The matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) MS analysis of compound 2 revealed the following: [M+H] calculated (monoisotopic)
15 1470.7, observed 1470.7.

Synthetic peptides 3 (SEQ ID NO: 10), 4 (SEQ ID NO: 11), and 5 (Figure 5A) were synthesized using established peptide synthesis protocols. Peptide 3 contained both the Gcn4 dimerization domain and the AH activation domain at the carboxy-terminus of the peptide. Peptide 4 contained the dimerization domain from
20 Gcn4 without the AH peptide; and peptide 5 contained the AH domain alone. In addition, each of peptides 3, 4, and 5 when synthesized included an amino-terminal Cys residue to facilitate linkage to compound 2. In separate reactions, polyamide 2 (1 μmol) was combined with either peptide 3, 4, or 5, (0.8-1 μmol) in 5% 1-methyl-2-pyrrolidinone in 6 M Gn HCl, 100 mM potassium phosphate buffer (pH 7.3), and
25 10% (vol/vol) thiophenol was added to this solution. Reaction progress was monitored by analytical HPLC and upon completion (3-5 days), purification of the mixture by reversed-phase HPLC resulted in isolation of the desired conjugates. Yields and characterization were as follows: compound 3 (PA-Gcn4-AH): 11%, MALDI-TOF [M+H] (average mass) calculated 7465.4, observed 7465.4;
30 compound 4 (PA-Gcn4): 21%, MALDI-TOF: [M+H] (average mass) calculated

SUBSTITUTE SHEET (RULE 26)

5159.8, observed 5159.9; compound 5 (PA-AH): 22%, MALDI-TOF: [M+H]
(average mass) calculated 3774.2, observed 3774.5.

Synthesis of Conjugates 8 and 9 (Figure 6). The ethylene glycol-derived linker was prepared as the *N*-*t*-butoxycarbonyl amino acid for use in solid-phase
5 synthesis from 4,7,10-trioxa-1,13-tridecanediamine by monoprotection with *N*-*t*-
butoxycarbonyl anhydride followed by reaction with diglycolic anhydride and
incorporated into polyamides 6 and 7 according to established protocols. Baird et al.
(1996). Transformation into conjugates 8 (PA-1L-AH) and 9 (PA-2L-AH) was
accomplished as described above. Yields and characterization: compound 8: 12%,
10 MALDI-TOF MS [M+H] (average mass) calculated 4164.7; observed 4164.6;
compound 9: 11%, MALDI-TOF MS [M+H] (average mass) calculated 4482.0,
observed 4482.2.

DNase I Footprinting Titrations. Quantitative DNase I footprinting
titrations were carried out in accordance with established protocols on a 3-³²P-
15 labeled 271-bp pPT7 *Eco*RI/*Pvu*II restriction fragment. Trauger et al. supra; and
Senear et al. (1986) *Biochem.* 25:7344-7354.

In Vitro Transcription Assays. The template plasmid was constructed by
closing a 78-bp oligomer bearing three cognate palindromic sequences into a *Bgl*II
site 30 bp upstream of the TATA box of pML 53. This plasmid has the AdML
20 TATA box 30 bp upstream of a 277-bp G-less cassette. The "mismatch" template
(containing a substitution of a T/A base pair with a G/C base pair, as compared to
the "match," or target site) was constructed by cloning a 78-bp oligomer containing
three palindromic "mismatch" sites into a *Bgl*II site 30 bp upstream of the TATA
box of pML 53. Gu et al. (1999) *Mol. Cell.* 3:97-108. For each reaction, 20 ng of
25 plasmid (30 fmol of palindromic sites) was preincubated with conjugate for 75 min
before the addition of 90 ng of yeast nuclear extract in a 25 μ L reaction volume
under standard conditions. Lue et al. (1991) *Met. Enzymol.* 194:545-55; and Lue et
al. (1989) *Science* 246:661-664. The reactions were processed as described (Lue et
al. (1991); and Lue et al. (1989)) and resolved on 8% 30:1 polyacrylamide gels
30 containing 8 M urea. Gels were dried and exposed to photostimulatable
phosphorimaging plates (Fuji). Data were visualized by using a Fuji
PhosphorImager followed by quantification using MACBAS software (Fuji).

Results and Discussion

Synthesis of Artificial Activators. The hairpin polyamides and peptides were synthesized by solid-phase protocols, and the peptides each contained an N-terminal cysteine for subsequent attachment to the polyamide. Polyamide 1 then
5 was treated with thiolane-2,5-dione followed by benzyl bromide to produce thioester 2 functionalized for use in the native ligation procedure described by Kent and colleagues (Fig. 5). Dawson et al. (1994). Three polyamide-peptide conjugates were prepared by this sequence. Conjugate 3 (PA-Gcn4-AH) contained the eight-ring hairpin polyamide as the DNA binding module in addition to residues 251-281
10 of the yeast protein Gcn4 as a dimerization element and the designed peptide AH as the activating region. The two control compounds 4 (PA-Gcn4) and 5 (PA-AH) each lacked one or another of components critical for activator function.

Based on available crystal structures (Ellenberger et al. (1992); and Dawson et al. (1994)), it was expected that the target site specificity of the polyamide would
15 target the synthetic regulatory compounds to the palindromic binding site, as shown in Fig 7. The data from a quantitative DNase I footprinting titration between conjugate 3 and DNA containing the 19-bp site were fit by a cooperative binding isotherm (K_D 11 nM) (Fig. 8A) (Senear et al. supra; and Brenowitz et al. (1986) Met. Enzymol, 130:132-181), confirming this expectation. The decrease in overall
20 binding affinity of conjugate 3 relative to the parent hairpin polyamide can be attributed to the attachment of the peptide at the C-terminal position of the polyamide, known to have a deleterious effect.

Activation of Transcription. Conjugate 3 (PA-Gcn4-AH) activated transcription in yeast nuclear extracts on a DNA template containing three
25 palindromic binding sites upstream of the start site (Fig. 8B). Lue (1991); and Lue (1989). Thus, inclusion of compound 3 at 200 nM concentration in the reactions resulted in 13-fold levels of activated transcription over basal levels. In control experiments, polyamide 1 alone or polyamide coupled to the Gcn4 dimerization domain (polyamide 4) but lacking AH did not stimulate transcription. Furthermore,
30 activation depended on the presence of cognate polyamide binding sites upstream of the transcription initiation site. Thus, on a template with palindromic sites

containing a single base pair mismatch at each half site, compound 3 failed to significantly activate transcription (Fig. 8C).

Time Dependence of Activation. The time course experiment (the results of which appear in Figure 9A) revealed that the activation profile of conjugate 3 (PA-Gcn4-AH) was consistent with that previously determined for protein transcriptional activators. Wootner et al. (1990) *J. Biol. Chem.* 265:8979-8982. At 20 min., the level of transcription was 40-fold above the basal level. Figure 9B shows that at high concentrations of free AH peptide, the activation elicited by compound 3 was decreased by about 50%. AH peptide thus competes with the DNA-bound compound 3 for binding to the transcriptional machinery in a phenomenon referred to as squelching. Gill et al. (1988) *Nature* 334:721-724; and Tasset et al. (1990) *Cell* 62:1177-1187. This demonstrates that DNA-tethered AH recruits the transcriptional machinery to the nearby promoter.

Functional Role of Dimerization Element. The functional necessity of a dimerization element was investigated by the evaluation of conjugates containing the activator peptide AH separated by flexible straight-chain linkers of 12 atoms (compound 5), 36 atoms (compound 8), and 55 atoms (compound 9) (Fig. 6). As shown in Figure 10A, compound 8 (PA-1L-AH) activated transcription at approximately 50% the level of compound 3 (PA-Gcn4-AH). Increasing the linker length to 55 atoms (compound 9) did not result in a further increase in activation levels; this is likely because of the flexibility of the linker moiety, which can not project AH fully away from DNA. The use of a shorter linker (compound 5, PA-AH) provided a compound that activated transcription 25% as well as conjugate 3 (PA-Gcn4-AH), confirming that spatial separation of the activator moiety from DNA plays a role in the efficiency of activation.

Two observations also demonstrate that compounds 5, 8, and 9 do not dimerize. As shown in Figure 10B, data from quantitative DNase I footprinting titrations were fit by noncooperative isotherms (K_D for compound 5 = 19 nM; K_D for compound 8 = 32 nM). Furthermore, in contrast to titrations containing compound 3 (PA-Gcn4-AH), DNase I-mediated cleavage was observed at positions between the monomeric binding sites within the palindrome.

These data demonstrate that synthetic regulatory compounds according to the invention can regulate transcription of a desired gene. These results also show that dimerization per se between two compounds is not required for activator function, nor is a natural DNA binding domain essential for activation.

5

EXAMPLE 6

Altering size, composition, site of attachment
of activating regions on the polyamide

As described above, naturally occurring transcriptional activator proteins minimally comprise two functions, one for DNA binding and the other for
10 activation. Example 5 describes the construction and testing of synthetic regulatory compounds that comprise non-natural nucleic acid binding moieties and various linkers in combination with the designed amphipathic AH activator peptide, a motif later found to occur in many native activation proteins. One of those synthetic regulatory activators, 4.2 kDa in size, comprised an eight-ring DNA binding hairpin
15 polyamide tethered through a flexible ethylene glycol-based polyether linker to the 20 residue AH activating peptide stimulated high levels of promoter-specific transcription. One surprise from that study was that a dimerization domain was unnecessary for function of the activation moiety.

This example further demonstrates the utility of the nucleic acid binding
20 moiety-linker-regulatory moiety motif by showing that even smaller synthetic ligands that mimic the activities of naturally occurring regulatory proteins can be successfully assembled and have regulatory function. Small synthetic regulatory compounds are preferred in order to increase the probability of membrane permeability without appreciable loss of specific gene-regulating activity.

25 In this example, synthetic regulatory compounds are provided that comprise sequence-specific polyamides attached via a linker to an activation peptide comprised of 8 or 16 residues derived from the activation domain of the potent viral activator VP16. The 16 residue activation moiety coupled to the polyamide via a linker activated transcription three fold better than the analogous AH conjugate
30 described in Example 5. Altering the site-of-attachment of the activation moiety on the polyamide allowed reduction of the intervening linker from 36 atoms to eight without significant diminution of the activation potential. Also provided are

synthetic regulatory compounds containing different polyamides to target a different sequence without compromising activator function, further emphasizing the generality of this motif.

These synthetic regulatory compounds have tunable potency effected by the size and identity of the activating moiety as well as the site of attachment to the polyamide. In particular, these results reveal a potency two to three times greater than PA-1L-AH conjugate 8 in Example 5) with concomitant decreases in molecular weight of 21% and 12%, respectively, and that these synthetic polyamide-linker-activator regulatory compounds bind to their cognate target sequences upstream of the AdML promoter. These results also show that levels of compounds required for full promoter occupancy correspond to the levels required to elicit maximal activation in vitro, and further confirm that changing the DNA binding domain can be accomplished while retaining significant transcription-activating function.

Materials and Methods

Referring to Figure 11, conjugates 1, 5, and 10 were synthesized and transformed. Polyamides 1, 5, and 10 were prepared according to established solid phase synthesis protocols and subsequently transformed into conjugates 2, 3, 4, 6, 7, 8, and 11 by the previously reported methods. Conjugate 9 was prepared in an analogous fashion. The identity of all conjugates was verified by MALDI-TOF mass spectrometry.

Characterization: conjugate 2 (PA-1L-AH): MALDI-TOF [M+H] (average mass) calculated 4164.7, observed 4164.6; conjugate 3 (PA-1L-VP2): MALDI-TOF: [M-H] (average mass) calculated 3670.1, observed 3670.1; conjugate 4 (PA-1L-VP1) MALDI-TOF [M+H] (monoisotopic mass) calculated 2763.25, observed 2763.53; conjugate 6 (PA-AH): MALDI-TOF: [M+H] (average mass) calculated 3774.2, observed 3774.5; conjugate 7 (PA-VP2): MALDI-TOF [M+H] (monoisotopic mass) calculated 3280.4, observed 3280.5; conjugate 8 (PA-VP1): MALDI-TOF [M+H] (monoisotopic mass) calculated 2374.0, observed 2374.0; conjugate 9 (PA-(py)-VP2): MALDI-TOF [M+H] (monoisotopic mass) calculated 3280.4, observed 3280.7; conjugate 11: MALDI-TOF [M+H] (monoisotopic mass) calculated 3670.6, observed 3670.7.

DNase I Footprinting Titrations. A 363-bp 5'-³²P-labeled PCR fragment was generated from template plasmid pAZA812 in accordance with standard protocols and isolated by nondenaturing gel electrophoresis. All DNase I footprinting reactions were carried out in a volume of 40 μ L. 0.8 ng/ μ L of plasmid pPT7 was used in these reactions as unlabeled carrier DNA. A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were 50 mM HEPES, 100 mM KOAc, 15 mM Mg(OAc)₂, 5 mM CaCl₂, 6.5% glycerol, 1 mM DTT, pH 7.0 and 15 kcpm 5'-radiolabeled DNA. The solutions were equilibrated for 75 min. at 22°C. Cleavage was initiated by the addition of 4 μ L of a DNase I stock solution and was allowed to proceed for 7 min. at 22°C. The reactions were stopped by adding 10 μ L of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL glycogen, and 30 μ M base pair calf thymus DNA and then ethanol-precipitated. The cleavage products were resuspended in 100 mM Tris-borate-EDTA/80% formamide loading buffer, denatured at 85°C for 10 min, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% cross-link, 7 M urea) at 2000 V for 2 h. 15 min. The gels were dried under vacuum at 80°C and quantitated using storage phosphor technology.

In Vitro Transcription Assays. Template plasmid pAZA812 was constructed by cloning a 78 bp oligomer bearing three cognate palindromic sequences for conjugates 2, 3, 4, 6, 7, 8, and 9 into a *Bgl*III site 30 bp upstream of the TATA box of pML Δ 53. This plasmid has the AdML TATA box 30 bp upstream of a 277 bp G-less cassette. Template pAZA813 was constructed by cloning a 78 bp oligomer containing three palindromic cognate sites for compound 11 into a *Bgl*III site 30 bp upstream of the TATA box of pML Δ 53 (22). For each reaction, 20 ng of plasmid (30 femtomoles of palindromic sites) was preincubated with conjugate for 75 minutes prior to the addition of 90 ng of yeast nuclear extract in a 25 μ L reaction volume under standard conditions. The reactions were processed as previously described and resolved on 8% 30:1 polyacrylamide gels containing 8 M urea (see Example 5). Gels were dried and exposed to photostimulatable phosphorimaging plates (Fuji Photo Film Co.). Data were visualized using a Fuji phosphorimager followed by quantitation using MacBAS software (Fuji Photo Film Co.).

SUBSTITUTE SHEET (RULE 26)

Results and Discussion

Synthesis of Synthetic Activators. The activation domain (residues 411-490) of VP16 fused to a DNA binding protein yielded a synthetic regulatory compound that activated transcription with a potency comparable to strong natural
5 activator proteins. Dissection of this and other activation domains has identified minimal units that activate transcription, these modules are often surreptitiously iterated in natural activating regions. In VP16, one such minimal moiety comprises eight amino acids. When iterated, the activation potential of the consequent peptide increased in a synergistic rather than an additive manner. This property of activating
10 regions was adopted in order to design artificial activators of varying strengths. A series of polyamide-linker-(VP16 minimal module)_x compounds (where x is 1 or 2) was designed, each of which comprises: a hairpin polyamide designed to target the cognate sequence 5'-TGTTAT-3'; a flexible tether of varying length (12 or 36 atoms); and one of three different activating regions (see Figure 11). The three
15 activating regions were AH and one (VP1) or two (VP2) tandem repeats of the eight amino acid sequence derived from VP16 (SEQ ID NOs: 16 and 17).

Also, it was determined that a critical role of the linker is to facilitate projection of the activating moiety away from the DNA for productive interaction with the transcriptional machinery. Initially, this was achieved by conjugating a
20 polyether linker to the C-terminus of hairpin polyamides. Conjugation via linkage at an internal pyrrole residue also appeared attractive, as solution studies and x-ray crystallographic data demonstrated that the N-methyl group of the pyrrole residues is directed outward from the minor groove when a polyamide binding in the 2:1 motif forms a sequence-specific complex with dsDNA.

25 Compounds 1 and 5 and conjugates 2-4 and 6-9 were prepared by previously reported methods, and their identity was verified by matrix-assisted laser-desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. Conjugates 2, 3, and 4 included a 36 atom polyether tether as the linker. Conjugates 6, 7, and 8 utilized a shorter 12 atom linking region.

30 **Promoter occupancy under transcriptional conditions.** The dissociation constant (K_D) for compound 2 binding to its cognate site (5'-TGTTAT-3') was measured as 32 nM using quantitative DNase I footprinting titrations. However, the

conditions for those experiments were substantially different than those of the typical in vitro transcription assays. For example, in quantitative DNase I footprinting experiments, for accurate determination of K_D for the ligand - DNA complex the concentration of DNA preferably is at least about a 10-fold excess of polyamide relative to DNA. In contrast, in vitro transcription assays employ a relatively high concentration of plasmid DNA (0.8 ng/ μ L), which can cause a decrease in the occupancy of a target nucleotide sequence by 10- to 100-fold.

To investigate the binding behavior of compound 2 under such conditions, a 5'-³²P-labeled 363 bp DNA fragment containing both the promoter region and 140 bp of the G-less cassette reporter was generated. To each footprinting reaction, unlabeled plasmid DNA was added to bring the total concentration of polyamide binding sites to a level equivalent to those utilized in transcription assays. As anticipated, DNase I footprinting titrations run under the conditions of the in vitro transcription assays revealed that 50% occupancy of the three dimeric binding sites occurs at a concentration of 215 nM, approximately a 7-fold increase over the measured K_D (Figure 12). Moreover, compound 2 binds specifically to the three palindromic sites with no binding observed at concentrations of up to 1 μ M in the G-less cassette region or at single base pair mismatch sites present elsewhere in the DNA fragment.

Promoter occupancy and the level of activation. An in vitro transcription titration experiment with compound 2 demonstrated that full occupancy of the promoter is necessary for maximal activator function are presented in Table 1.

Table 1

Conjugate	50% Occupancy
PA-1L-AH (2)	215 nM
PA-1L-VP2 (3)	210 nM
PA-1L-VP1 (4)	75 nM
PA(5)	15 nM
PA-VP2 (7)	75 nM
PA-VP1 (8)	110 nM
PA-(py)-VP2 (9)	175 nM

SUBSTITUTE SHEET (RULE 26)

At a concentration of 100 nM, detectable levels (>4-fold) of transcriptional activation were observed, and as the concentration was increased to provide full saturation of the dimeric sites, activation levels rose to 20-fold over-basal. These data show that optimal synthetic regulatory compound concentration for in vitro transcription assays can be predicted using data from DNase I footprinting titrations under such conditions. Therefore, additional DNase I footprinting titration experiments were carried out, revealing that the compounds used in this example required concentrations of 75 to 215 nM to attain 50% target site occupancy. Based upon these data, all subsequent in vitro transcription experiments were carried out at synthetic regulatory compound concentrations of 200-400 nM.

Activating potential corresponds to size and site-of-attachment of the activation moiety. The in vitro transcription experiments reveal that the activation strength of a synthetic regulatory compound is proportional to the size of the activating region, and that projection of the activating region away from the DNA enhances its functionality. When compared with compound 2 (PA-1L-AH), compound 8 (PA-VP1), which lacked a linker, is a poor activator, and that substituting VP1 with VP2 increased the activation strength of the resulting compound.

Consistent with the requirement for the activating moiety to access targets in the transcriptional machinery, projecting the VP1 or VP2 module via a longer linker further improved the activation potential the synthetic regulatory compounds. From Figure 12, it is apparent that a VP2 peptide attached to the C-terminus of a polyamide via a 36 atom linker (PA-1L-VP2; compound 3) was the most potent of the polyamide-linker-activating moiety compounds tested in this example. However, compound 9, where VP2 was attached to an internal pyrrole residue via an eight atom linker, activated transcription robustly despite the absence of a long linker, demonstrating that projection of the activating region from this position is particularly effective. Transcriptional stimulation by each compound was dependent on the presence of cognate binding sites upstream of the promoter. Thus, on a template bearing mismatched binding sites, compounds 3, 7, and 9 did not stimulate transcription effectively.

Replacing the DNA binding moiety alters specificity. To demonstrate the generality of synthetic regulatory compound motif of the invention, the identity of the DNA recognition domain was changed and did not compromise function. Specifically, compound 11 targeted the sequence 5'-AGGTCA-3', and incorporated the polyether linker as the linking domain and VP2 as the activating region. This compound had proven to be the most active of all tested in the experiments reported this example (Figure 13). Control polyamide 10 and compound 11 were prepared by established protocols. As shown in Figure 14, a template bearing three binding sites 40bp upstream of the AdML G-less cassette reporter was constructed. The substitution of the nucleic acid binding moiety resulted in, as expected, a synthetic activator that specifically targeted the template bearing its cognate DNA binding sites and not the template bearing sites for the previous polyamide (Figure 14).

EXAMPLE 8

Metazoan Cell Culture Experiments

This example describes experiments that can be used to assess whether a synthetic regulatory compound according to the invention is cell permeable, and cell permeability for members of this class of compounds is demonstrated for the first time.

The synthetic regulatory compounds used in this example employed polyamide molecules as non-natural nucleic acid binding moieties. The compounds were tested for cell permeability against two cell lines, SKOV (a cisplatin resistant human ovarian cancer cell line) and 293T (a human renal cell line) cells. The compounds were also tested for their ability to activate a transiently transfected reporter gene the expression of which could be up-regulated by activation of a promoter functionally associated with one DR5 target site. Other cells, including, without limitation, COS, Cho, Jurkat, and HeLa cells, are also suitable for use with routine modifications in the experiments reported herein.

The reporter gene carried by the cells comprised a minimal HSV TK promoter driving the expression of a luciferase gene. Promoter activity was regulated by a consensus DR5 site approximately 50 bp upstream of the TK promoter. The DR5 site comprises a direct repeat of two consensus hexameric sequences separated by five nucleotides, the identity of which was irrelevant. One

of the consensus hexameric nucleotide sequences in the DR5 site was 5'-AGGTCA-3', which also corresponds to an estrogen receptor half-site.

The polyamide moieties of the synthetic regulatory compounds used in this example targeted the nucleotide sequence 5'-WGGWCA-3', and were fused via a PEG linker at the C-terminal tail to either the L- or the D- form of a VP2 activating region (AM_V154 and AM_V155, respectively). Prior to testing these two conjugates on reporter cells, these compounds were tested for their ability to activate transcription in standard cell-free yeast system that comprised a reporter construct having three tandem 5'-AGGTCA-3' sites located 40 bp upstream of the AdML TATA:G-less cassette. The in vitro transcription experiments were performed as reported above. As expected, the AM_V151 polyamide alone did not stimulate transcription over basal conditions. The synthetic regulatory compound AM_V154 (PA-1L-VP2) activated transcription about 12 fold, and AM_V155 (PA-1L-D-VP2) activated transcription about ten-fold in the in vitro assays. The observed levels of activation were consistent with the fact that only three half-sites, rather than three complete palindromes, were used in the reporter construct employed in the in vitro transcription assays.

The cells used for the cell permeability and cell-based reporter assays were passaged twice and then grown overnight in 6-well plates to sub-confluency in CO₂ incubators. Two micrograms of DR5-Luc DNA and 0.5 micrograms of CMV-βgal were transfected in each of the wells. One well on each plate contained no reporter construct, and one contained a strong CMV promoter driving the Luc reporter gene to check for transfection efficiency. SKOV cells were transfected using DOTAP (a cationic lipid transfection system), and the 293 cells were transfected by the standard calcium phosphate precipitation method.

After DNA addition, the cells were incubated for 12 hours and then washed with fresh DMEM media containing 10% charcoal-stripped Fetal Bovine Serum (to remove ligands of nuclear receptors that might activate transcription of the reporter). Cells were then allowed to recover for 12 hours in 2 mL of DMEM+15% FBS (stripped). After washing out this media, cells were supplied with 1 mL of DMEM+10% FBS(stripped) containing AM_V151, AM_V154, or AM_V155 at 1 μM concentrations.

Luciferase activity was measured using standard luminomitor techniques. The results of these experiments appear in Tables 2 and 3.

Table 2
Activity in SKOV Cells

	Reporter	Conjugate	Luciferase activity	normalized activation
1	Cells alone		464	
2	CMV-Luc		859304	
3	DR5-Luc	None	828117	
4	DR5-Luc	AM _v 151	400439	1
5	DR5 - Luc	AM _v 154	570395	1.425
6	DR5 -Luc	AM _v 155	703533	1.75

5

Table 3
Activity in 293T Cells

	Reporter	Conjugate	Luciferase activity	normalized fold activation
1	Cells alone		507	
2	CMV-Luc		34878848	
3	DR5-Luc	none	3226708	
4	DR5-Luc	AM _v 151	3532228	1
5	DR5 - Luc	AM _v 154	6351673	1.8
6	DR5 -Luc	AM _v 155	19692352	5.5

It is important to note that of the two cells lines, it was clear that the 293T cells were more readily transfected, as the stronger CMV-Luc reporter construct resulted in approximately 10-fold more luciferase activity as compared the DR5-TK-Luc (compare row 2 and 3). Despite the low transfection efficiency of the calcium phosphate technique, sufficient numbers of reporter constructs were taken up. Significantly, the synthetic regulatory compounds were taken up by each of the cell types tested, and they stimulated luciferase expression. Surprisingly, compounds containing the D-enantiomer of the activator stimulated more reporter gene

expression than the activator made using amino acid forms found naturally in proteins. While not wishing to be bound to a particular theory, it is believed that activating regions typically are unstructured in solution, and they are fairly hydrophobic. Such peptides are typically rapidly degraded in cells. Thus, the higher activity of the D-form in activating reporting expression in 293 cells can reflect an ability to resist proteolysis.

The contents of the articles, patents, and patent applications, and all other documents and electronically available information mentioned or cited herein, are hereby incorporated by reference in their entirety to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference. Applicants reserve the right to physically incorporate into this application any and all materials and information from any such articles, patents, patent applications, or other documents.

The inventions illustratively described herein can suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising", "including", "containing", etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed can be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of the inventions disclosed herein.

The inventions have been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of these inventions. This includes the generic description of each invention with a proviso or negative limitation removing any subject matter from the

genus, regardless of whether or not the excised material is specifically recited herein.

Other embodiments are within the following claims. In addition, where features or aspects of an invention are described in terms of a Markush group, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

CLAIMS

1. A cell permeable synthetic regulatory compound comprising:
- (a) a non-natural nucleic acid binding moiety;
 - (b) a regulatory moiety; and
 - 5 (c) a linker connecting the non-natural nucleic acid binding element to the regulatory moiety, or a pharmaceutically acceptable salt of such synthetic regulatory compound.
2. The synthetic regulatory compound according to claim 1 wherein the non-natural nucleic acid binding moiety is a molecular that binds to double-stranded
- 10 DNA via hydrogen bonds, van der Waals forces and/or electrostatic interactions.
3. The synthetic regulatory compound according to claim 1 wherein the non-natural nucleic acid binding moiety is selected from the group consisting of a polyamide, a peptide nucleic acid, and an oligonucleotide.
4. The synthetic regulatory compound according to claim 1 wherein the non-
- 15 natural nucleic acid binding moiety comprises a structure
- $$-Q_1-Z_1-Q_2-Z_2-\dots-Q_m-Z_m-$$
- wherein each of Q_1, Q_2, \dots, Q_m is independently selected from a heteroaromatic moiety and $(CH_2)_p$, wherein p is an integer between 1 and 3, inclusive; wherein each of Z_1, Z_2, \dots, Z_m is independently selected from the group consisting of a covalent
- 20 bond and a linking group; and m is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or 16.
5. A synthetic regulatory compound according to claim 4 wherein, with respect to the non-natural nucleic acid binding moiety, at least one of Q_1, Q_2, \dots, Q_m is a heteroaromatic moiety.
- 25 6. A synthetic regulatory compound according to claim 5 wherein, with respect to the non-natural nucleic acid binding moiety, the heteroaromatic moiety is selected from the group consisting of substituted and unsubstituted imidazole and pyrrole moieties.

7. A synthetic regulatory compound according to claim 4 wherein, with respect to the non-natural nucleic acid binding moiety, at least one of Z_1, Z_2, \dots, Z_m is a linking group having 2, 3, 4, or 5 backbone atoms.
8. A synthetic regulatory compound according to claim 7 wherein, with respect to the non-natural nucleic acid binding moiety, each of Z_1, Z_2, \dots, Z_m is a carboxamide group.
9. A synthetic regulatory compound according to claim 5 wherein, with respect to the non-natural nucleic acid binding moiety, at least 60% of Q_1, Q_2, \dots, Q_m are heteroaromatic moieties independently selected from the group consisting of substituted and unsubstituted imidazole and pyrrole moieties, and wherein each of Z_1, Z_2, \dots, Z_m is a carboxamide group.
10. A synthetic regulatory compound according to claim 9 wherein the non-natural nucleic acid binding moiety is cyclic.
11. A synthetic regulatory compound according to claim 9 wherein the non-natural nucleic acid binding moiety is capable of forming an intermolecular 2:1 binding motif under physiological conditions in the presence of a double-stranded DNA molecule comprising a corresponding target sequence.
12. A synthetic regulatory compound according to claim 9 wherein the non-natural nucleic acid binding moiety is capable of forming an intramolecular 2:1 binding motif under physiological conditions in the presence of a double-stranded DNA molecule comprising a corresponding target sequence.
13. A synthetic regulatory compound according to claim 9 wherein the non-natural nucleic acid binding moiety has, under physiological conditions, a binding specificity for its corresponding target sequence in a double-stranded DNA of at least about two as compared to a mismatch target sequence.
14. A synthetic regulatory compound according to claim 9 wherein the non-natural nucleic acid binding moiety has, under physiological conditions, a binding

specificity for its corresponding target sequence in a double-stranded DNA of at least about ten as compared a mismatch target sequence.

15. A synthetic regulatory compound according to claim 9 wherein the non-natural nucleic acid binding moiety has at least about submicromolar binding affinity for its corresponding target sequence in a double-stranded DNA under physiological conditions.
16. A synthetic regulatory compound according to claim 9 wherein the non-natural nucleic acid binding moiety has at least about subnanomolar binding affinity for its corresponding target sequence in a double-stranded DNA under physiological conditions.
17. A synthetic regulatory compound according to claim 1 wherein the regulatory moiety is selected from the group consisting of a small organic molecule, a lipid, a peptide, a carbohydrate, a nucleic acid and a peptide nucleic acid.
18. A synthetic regulatory compound according to claim 1 wherein the regulatory moiety that decreases expression of a target gene.
19. A synthetic regulatory compound according to claim 1 wherein the regulatory moiety that increases expression of a target gene.
20. A synthetic regulatory compound according to claim 1 wherein the linker is covalently linked to each of the nucleic acid binding moiety and the regulatory moiety.
21. A synthetic regulatory compound according to claim 20 wherein the linker comprises from 1 to about 200 spacing moieties.
22. A synthetic regulatory compound according to claim 21 wherein at least one of the spacing moieties of the linker is $-(CH_2)-$.
23. A synthetic regulatory compound according to claim 21 wherein the linker is attached to a terminal moiety of the non-natural nucleic acid binding moiety.

24. A synthetic regulatory compound according to claim 21 wherein the linker is attached to an internal moiety of the non-natural nucleic acid binding moiety.
25. A synthetic regulatory compound according to claim 24 wherein the internal moiety of the non-natural nucleic acid binding moiety is selected from the group
- 5 consisting of a γ -aminobutyric acid, β -alanine, a substituted pyrrole, an unsubstituted pyrrole, a substituted imidazole, and an unsubstituted imidazole.
26. A composition comprising a synthetic regulatory compound according to claim 1 and a carrier.
27. A composition according to claim 26 wherein the carrier is a
- 10 pharmaceutically acceptable carrier.
28. A composition according to claim 26 that is a liquid composition.
29. A composition according to claim 26 that is a dry composition.
30. A cell containing a synthetic regulatory compound according to claim 1.
31. A cell according to claim 30 selected from the group consisting of an animal
- 15 cell and a plant cell.
32. A cell according to claim 30 selected from the group of cells consisting of bovine, canine, equine, feline, murine, ovine, porcine, and primate cells.
33. A cell according to claim 29 that is a human cell.
34. A cell according to claim 29 that is in vitro.
- 20 35. A cell according to claim 29 that is in vivo.
36. A complex comprising a synthetic regulatory compound according to claim 1 complexed with a double-stranded DNA.

37. A method of forming a complex according to claim 36, comprising exposing a composition containing a double-stranded DNA comprising a target sequence to the synthetic regulatory compound.
38. A method according to claim 37 wherein the composition ~~is~~ a cell.
- 5 39. A method of regulating transcription of a regulatable gene, comprising exposing a double-stranded DNA encoding the regulatable gene to a synthetic regulatory compound according to claim 1 capable of regulating transcription thereof under transcription conditions.
40. A method according to claim 39 performed in vitro.
- 10 41. A method according to claim 39 performed in vivo.
42. A method of screening for a synthetic regulatory compound according to claim 1, comprising exposing, under transcription conditions, a double-stranded DNA encoding a regulatable gene to a plurality of test compounds, each of which comprises:
- 15 (a) a non-natural nucleic acid binding moiety targeted to a transcription-associated regulatory element of the regulatable gene;
- (b) an activation element; and
- (c) a linker connecting the non-natural nucleic acid binding element to the activation element,
- 20 and determining whether any of the test compounds regulates expression of the regulatable gene.
43. A method according to claim 42 performed in vitro.
44. A method according to claim 42 performed in vivo.
45. A method according to claim 42 wherein the regulatable gene is a marker
- 25 gene.
46. A synthetic regulatory compound comprising:

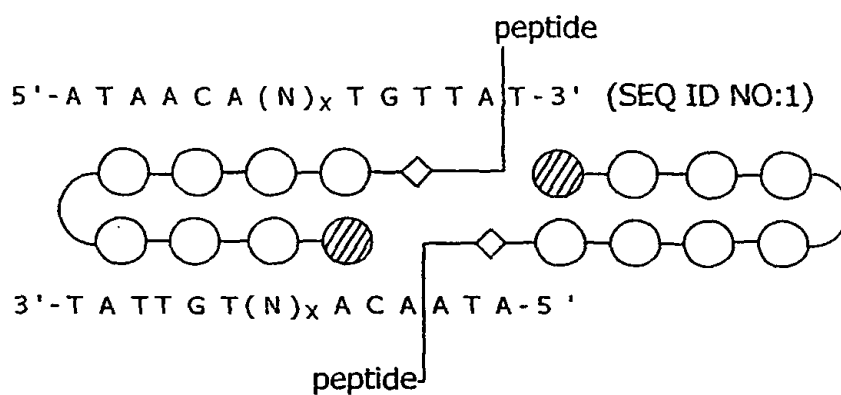
- (a) a first nucleic acid binding moiety and a second nucleic acid binding moiety, wherein at least one of the first or second nucleic acid binding moieties is a non-natural nucleic acid binding moiety;
- (b) a regulatory moiety; and
- 5 (c) at least one linker connecting the first or the second nucleic acid binding moiety to the regulatory moiety,
- or a pharmaceutically acceptable salt of such synthetic regulatory compound.
47. A synthetic regulatory compound comprising:
- (a) a non-natural nucleic acid binding moiety;
- 10 (b) a plurality of regulatory moieties; and
- (c) a linker connecting the non-natural nucleic acid binding moiety to the regulatory moieties,
- or a pharmaceutically acceptable salt of such synthetic regulatory compound.
48. A synthetic regulatory compound comprising:
- 15 (a) a non-natural nucleic acid binding moiety other than a hairpin polyamide;
- (b) a regulatory moiety; and
- (c) a linker connecting the non-natural nucleic acid binding moiety to the regulatory moiety,
- 20 or a pharmaceutically acceptable salt of such synthetic regulatory compound.
49. A synthetic regulatory compound comprising:
- (a) a non-natural nucleic acid binding moiety;
- (b) a regulatory moiety other than one that solely recruits a mediator complex; and
- 25 (c) a linker connecting the non-natural nucleic acid binding moiety to the regulatory moiety,
- or a pharmaceutically acceptable salt of such synthetic regulatory compound.
50. A synthetic regulatory compound comprising:
- (a) a non-natural nucleic acid binding moiety;

- (b) a regulatory moiety other than a small molecule; and
 - (c) a linker connecting the non-natural nucleic acid binding moiety to the regulatory moiety,
- or a pharmaceutically acceptable salt of such synthetic regulatory compound.

- 5 51. The synthetic regulatory compound according to claim 1 wherein the regulatory moiety targets chromatin remodeling activities.

1/17

A



B

Peptides:

7: C+YLLPTCIP (SEQ ID NO:2)

8: C+ linker + YLLPTCIP

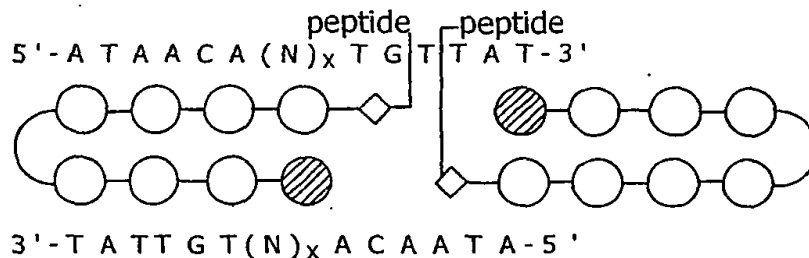
9: C+PRELDMLKMDSLQDIKALLTGLFVQD+YLLPTCIP (SEQ ID NO: 3)

SUBSTITUTE SHEET (RULE 26)

2/17

A

C-terminus Substitution

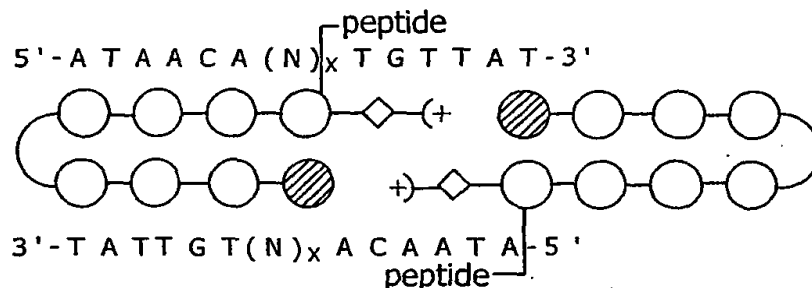


Peptides:

- 10: C + Gcn 4 (251-281) + LTGLFVQDYLLPTCIP (SEQ ID NO:4)
 11: C + Gcn 4 (251-281) + PEFPGIELQELQELQALLQQ (SEQ ID NO:5)
 12: C + Gal 4 (73-100) + PEFPGIELQELQELQALLQQ

B

Pyrrole Substitution

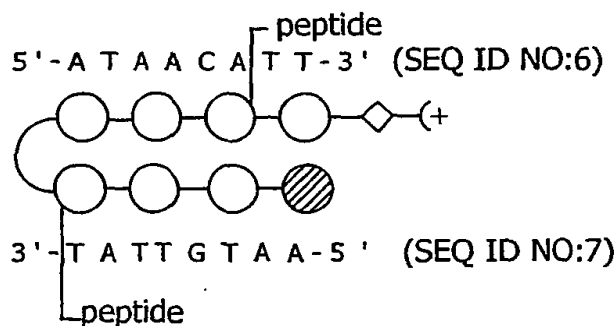


Peptides:

- 13: C + Gcn 4 (251-281) + LTGLFVQDYLLPTCIP
 14: C + Gcn 4 (251-281) + PEFPGIELQELQELQALLQQ
 15: C + Gal 4 (73-100) + YLLPTCIP
 16: C + Gal 4 (73-100) + PEFPGIELQELQELQALLQQ
 17: C + linker + YLLPTCIP

C

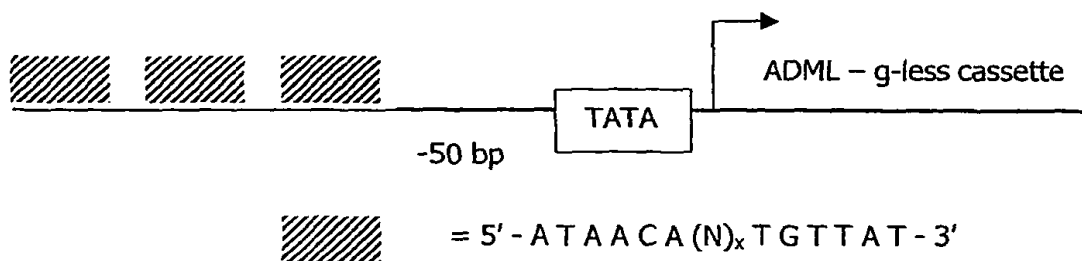
Intramolecular Dimerization



Peptides:

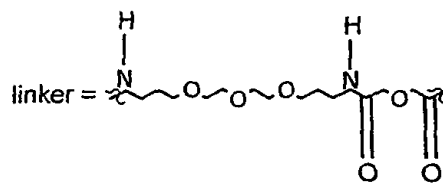
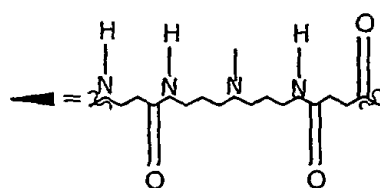
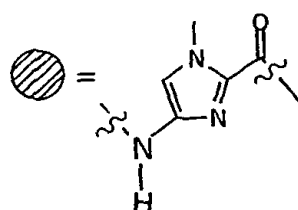
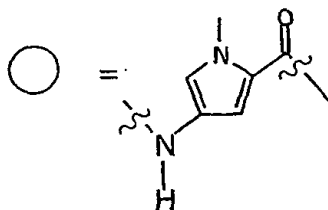
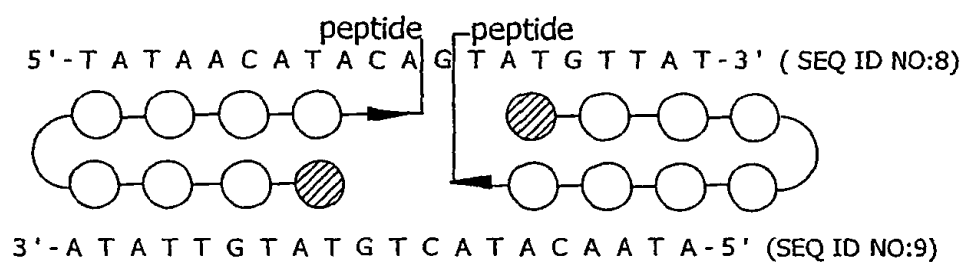
- 18: C + linker + YLLPTCIP
 19: C + Gal 4 (73-100) + YLLPTCIP
 20: C + Gal 4 (73-100) + PEFPGIELQELQELQALLQQ

3/17

 $X = 3, 5, \text{ or } 7$

ADML: adenovirus major late promotor

4/17

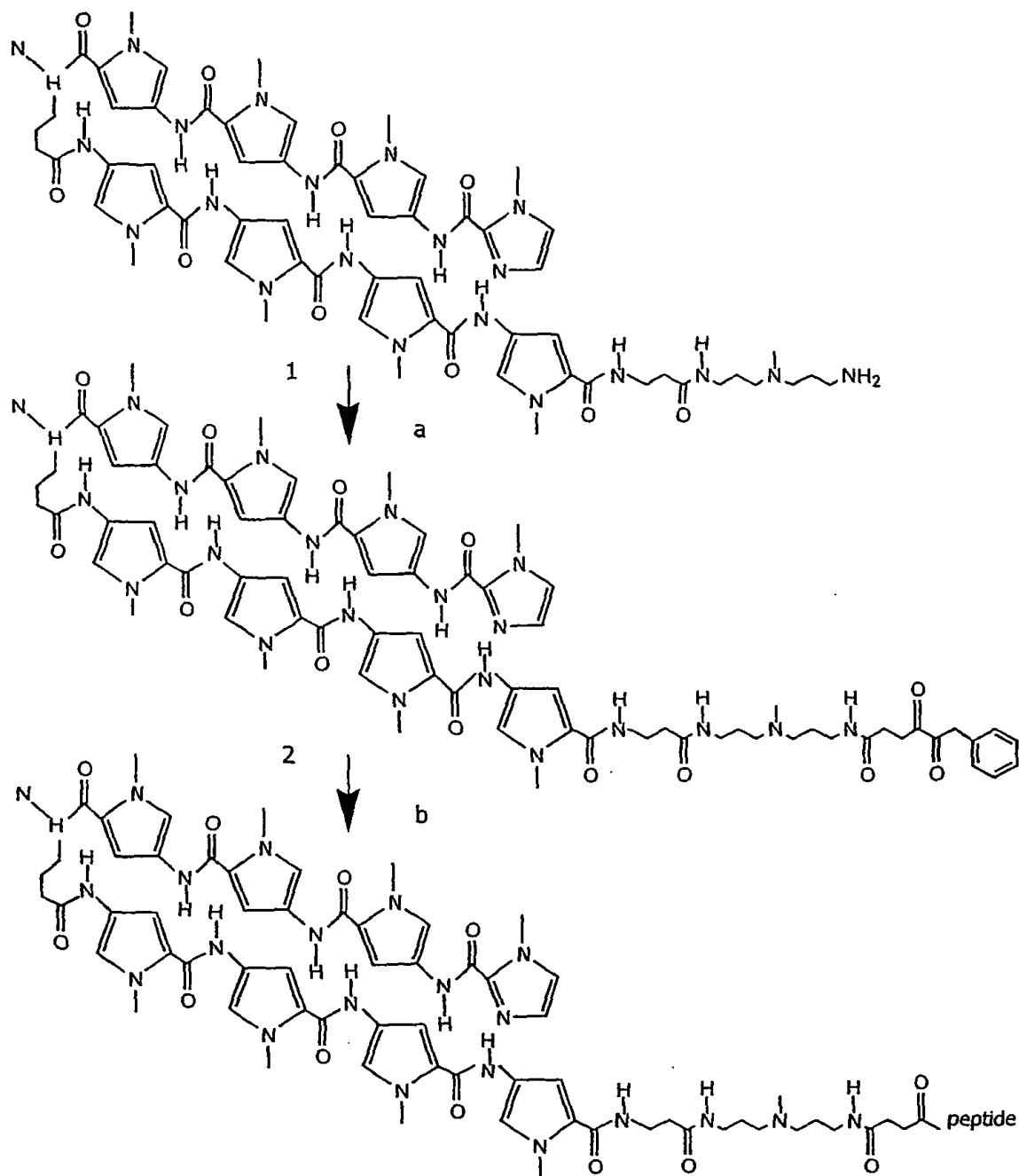


Peptides:

- 11: C + Gcn 4 (251-281) + P E F P G I E L Q E L Q E L Q A L L Q Q
- 21: C + scrambled Gcn 4 (251-281) + P E F P G I E L Q E L Q E L Q A L L Q Q
- 22: linker + C P E F P G I E L Q E L Q E L Q A L L Q Q
- 23: 2 x linker + C P E F P G I E L Q E L Q E L Q A L L Q Q
- 24: C + P E F P G I E L Q E L Q E L Q A L L Q Q
- 25: C + Gcn 4 (251-281)
- 26: linker
- 27: 2 x linker
- 28: polyamide only

SUBSTITUTE SHEET (RULE 26)

5/17



Peptides:

3 C - KQLEDKVEELLSKNYHLENEVARLKKLVGER-PEFPGIELQELQELQALLQQ (SEQ ID NO:10)

Gcn4 (251-281)

AH

4 C-KQLEDKVEELLSKNYHLENEVARLKKLVGER (SEQ ID NO:11)

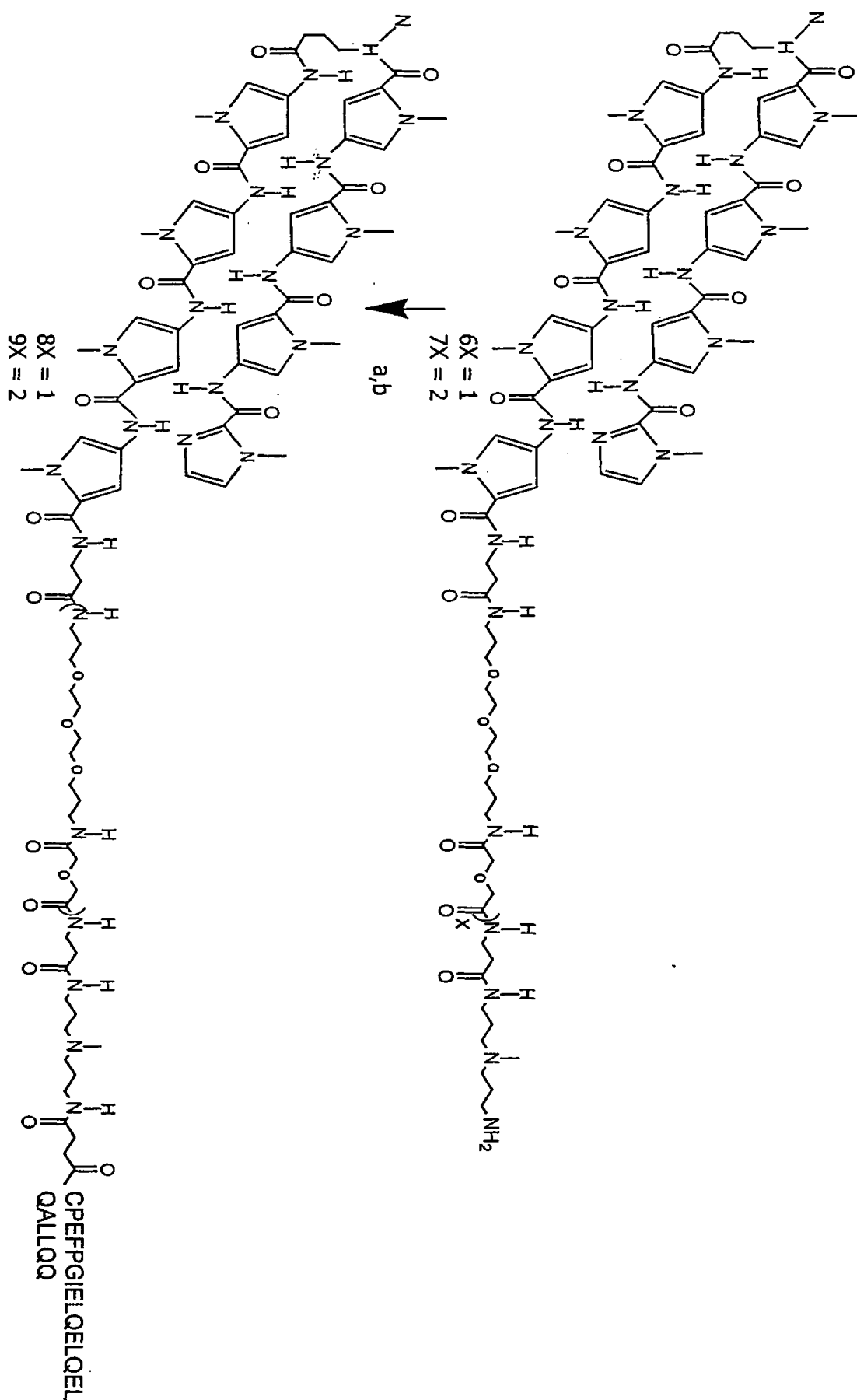
Gcn4 (251-281)

5 C-PEFPQUELQELQELQALLQQ

AH

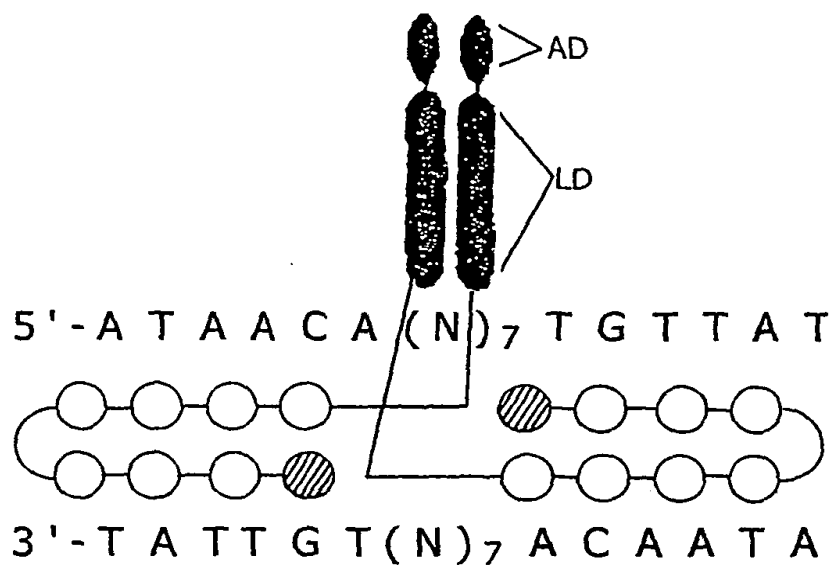
SUBSTITUTE SHEET (RULE 26)

6/17



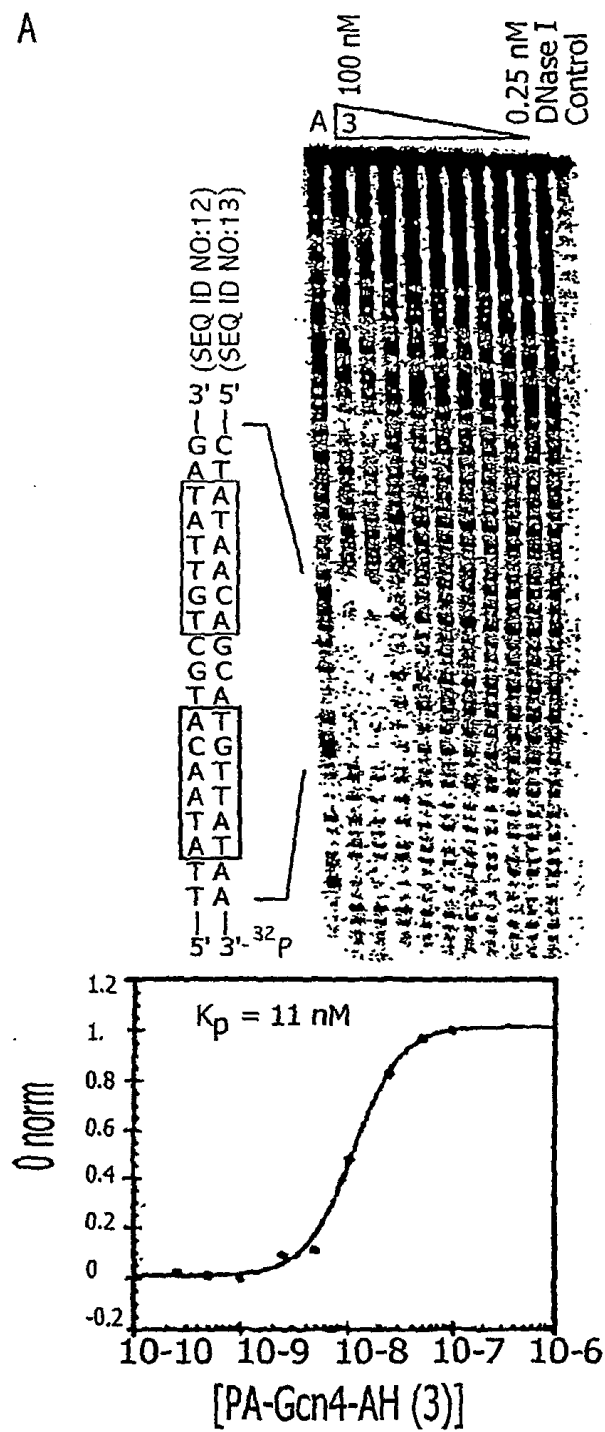
SUBSTITUTE SHEET (RULE 26)

7/17



8/17

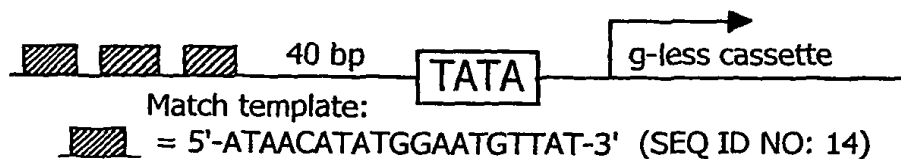
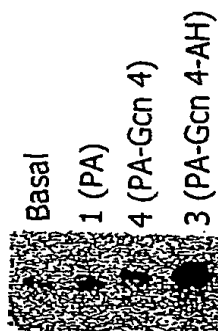
A



SUBSTITUTE SHEET (RULE 26)

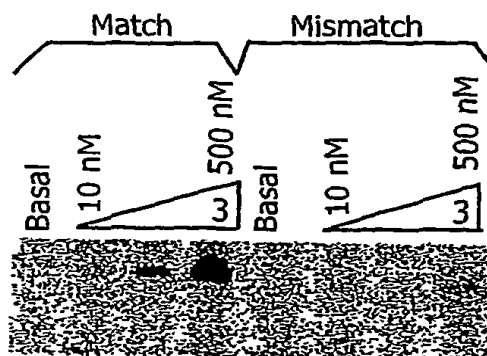
9/17

B



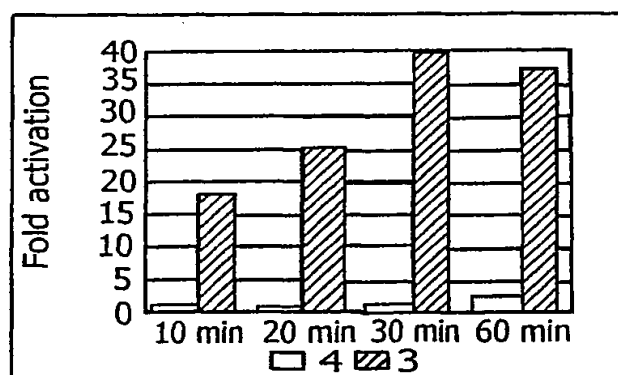
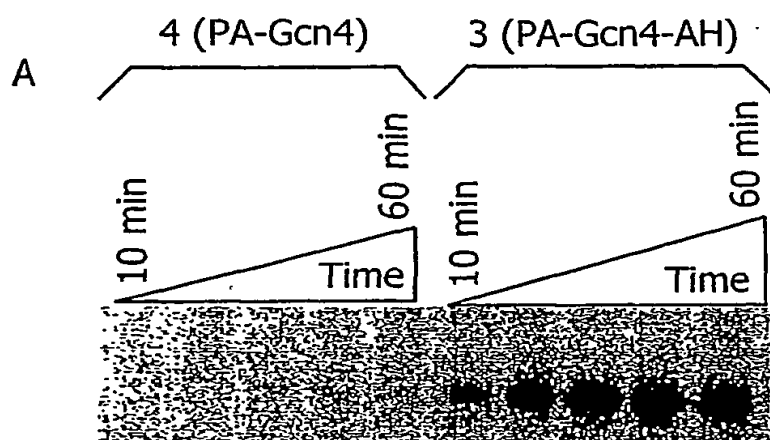
Mismatch template:
 = 5'-ATACCATATGGAATGGTAT-3' (SEQ ID NO: 15)

C

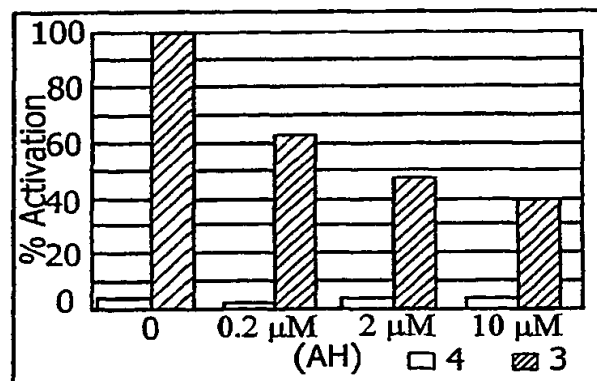
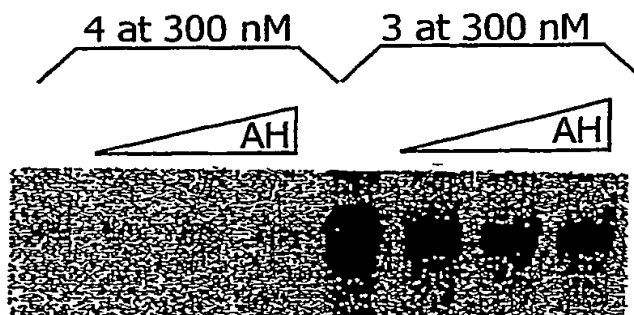


SUBSTITUTE SHEET (RULE 26)

10/17

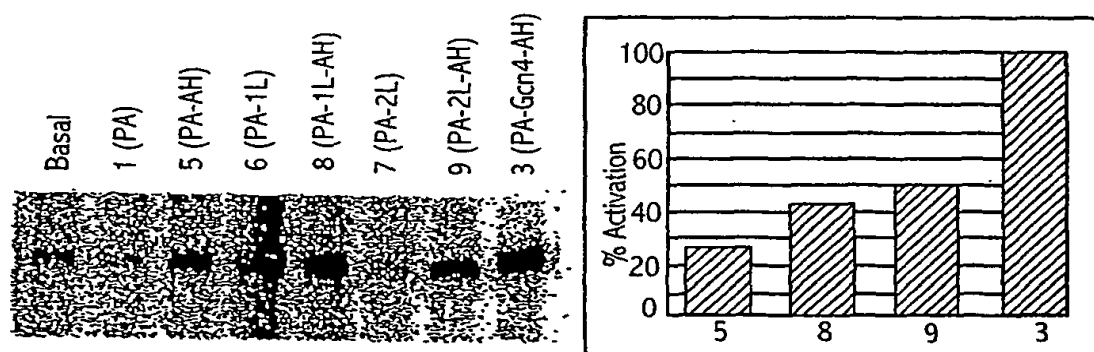


B

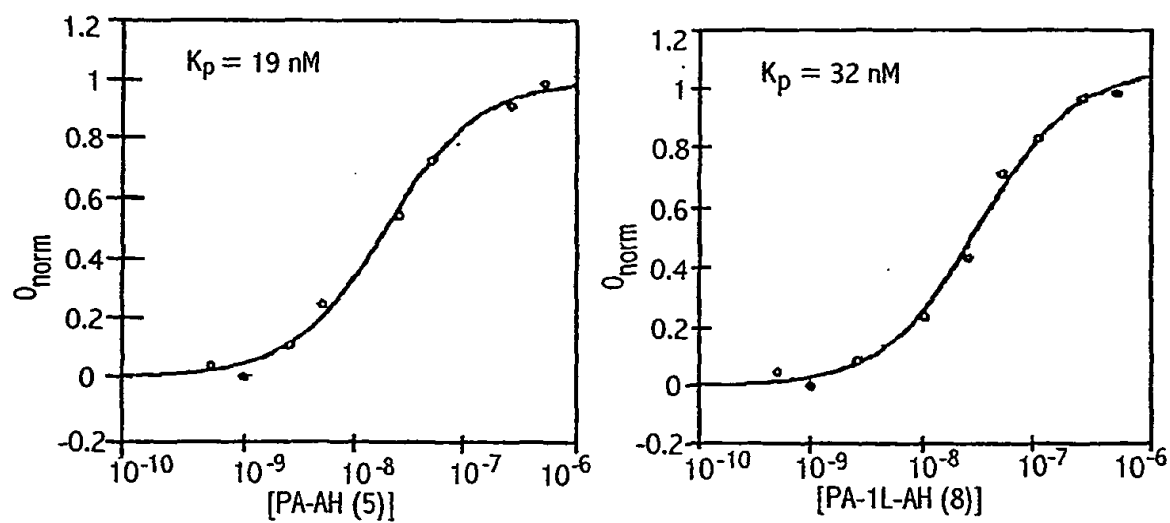


11/17

A

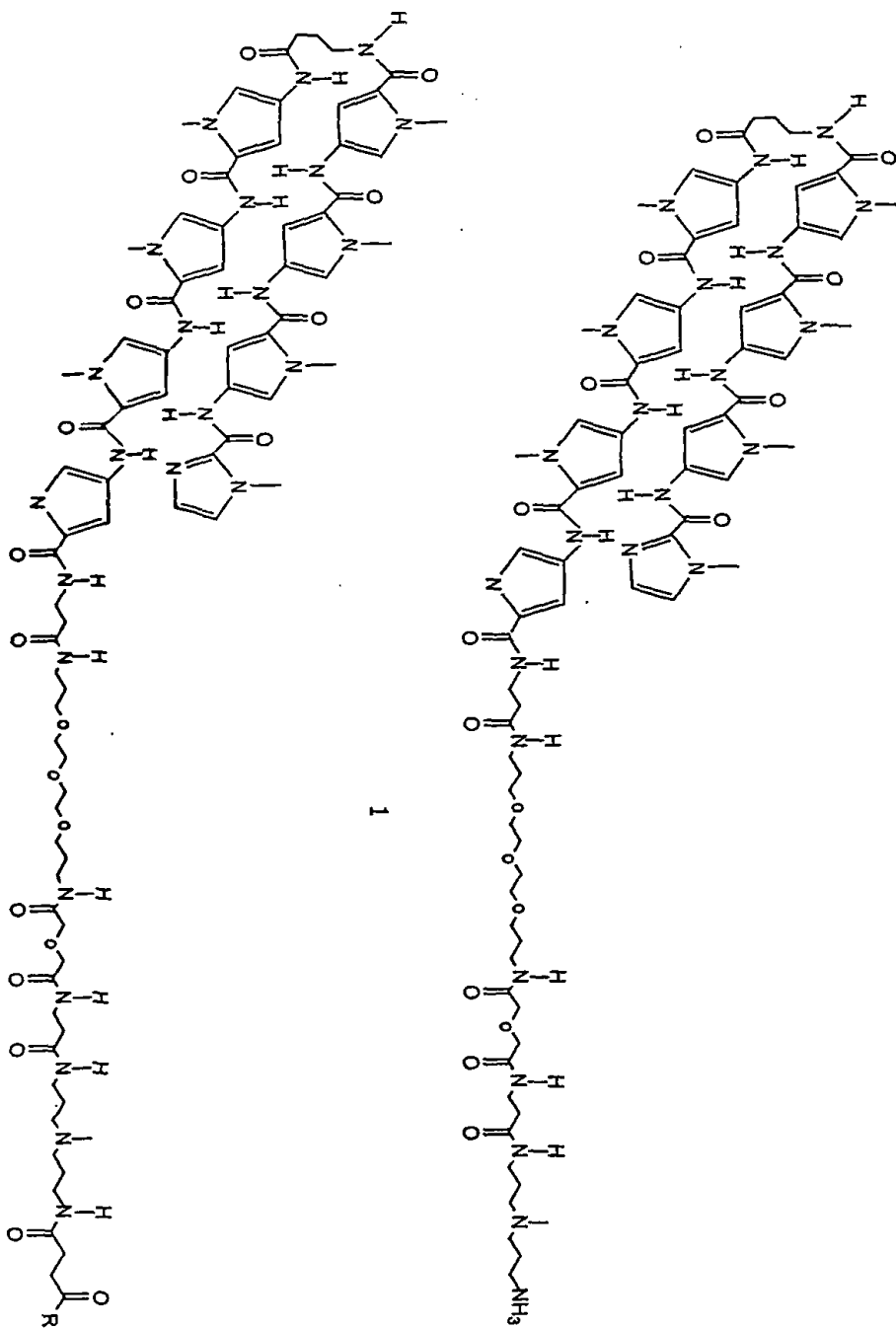


B



SUBSTITUTE SHEET (RULE 26)

12/17



R

2 C-PEFPGIELQELQELQALLQ

AH

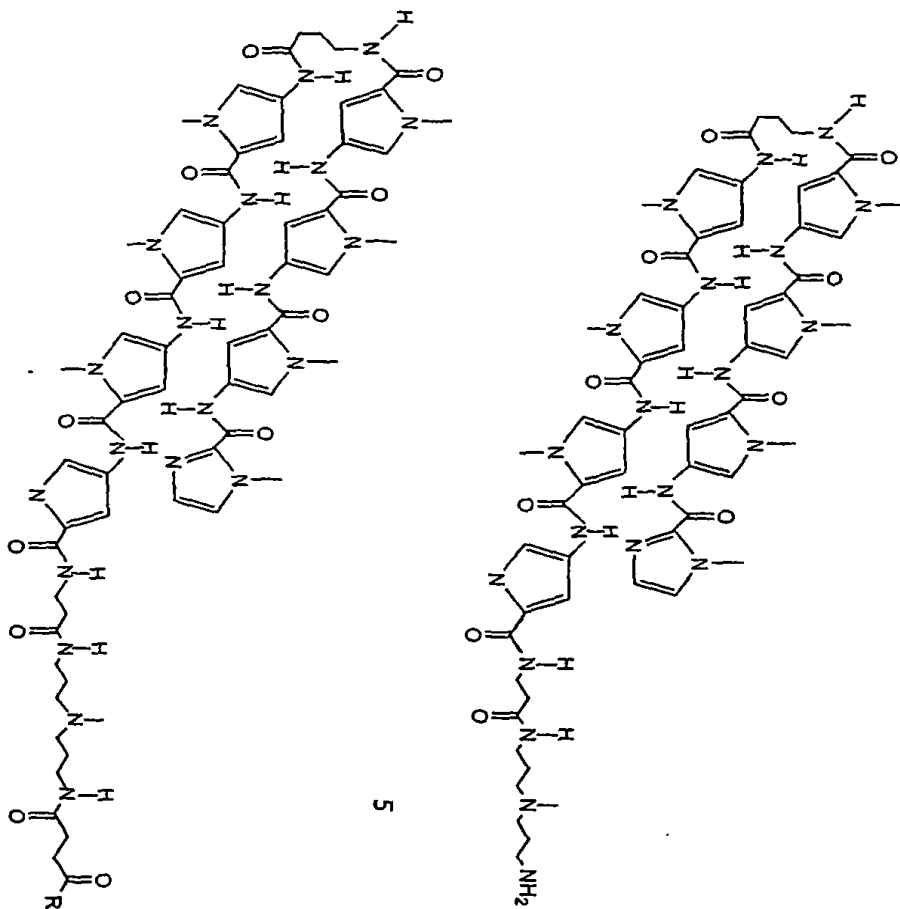
3 C-DFDLDMLGDFDLDMLG (SEQ ID NO: 16)

VP2

4 C-DFDLDMLG (SEQ ID NO: 17)

VP1

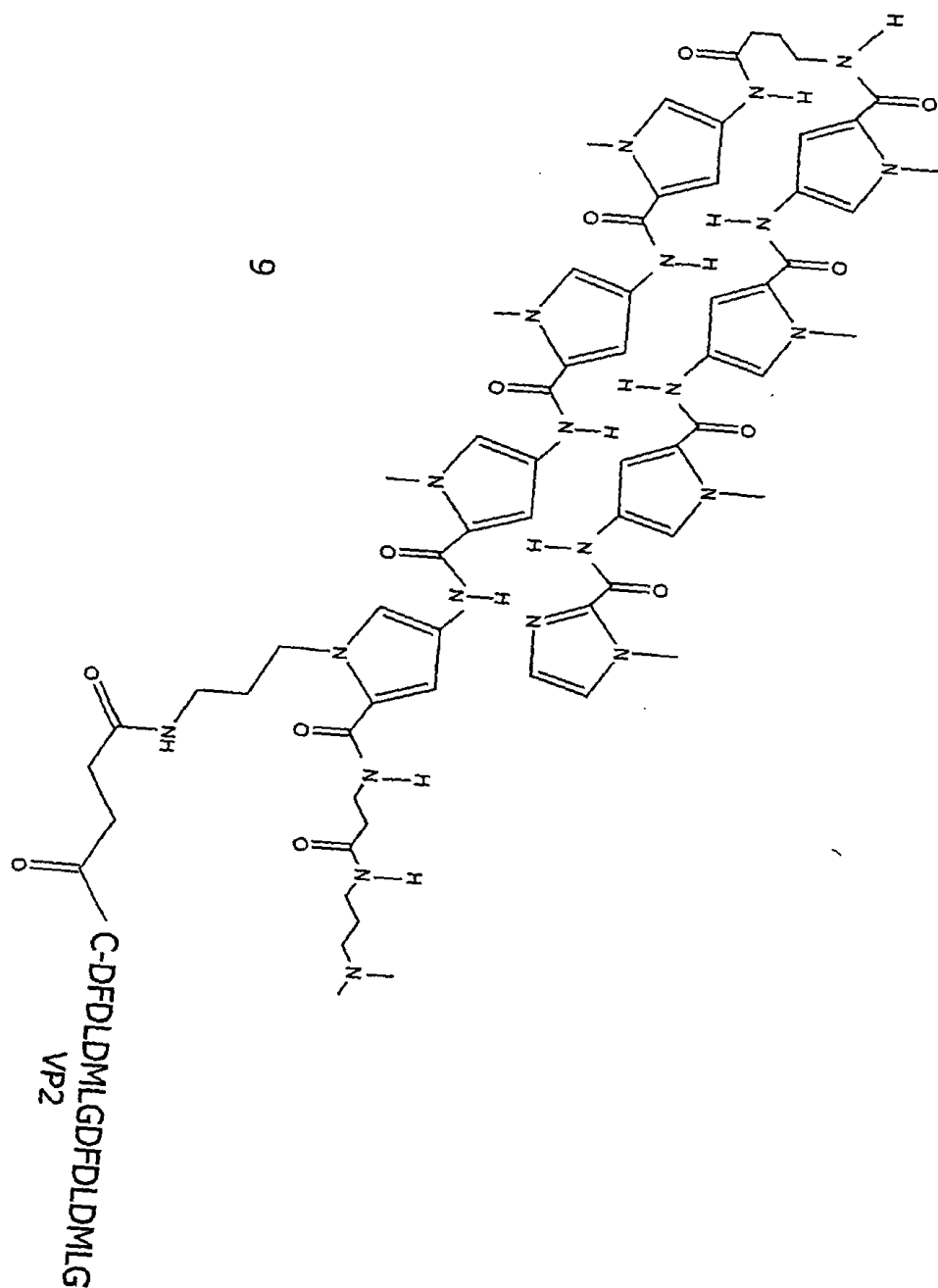
13/17



6 C-PEFPGIELQELQELQALLQ
 AH
 7 C-DFDLDMLGDFDLDMLG
 VP2
 8 C-DFDLDMLG
 VP1

SUBSTITUTE SHEET (RULE 26)

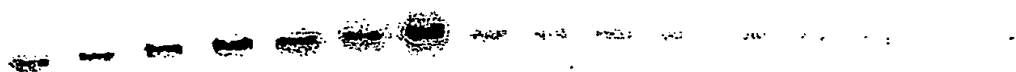
14/17



SUBSTITUTE SHEET (RULE 26)

15/17

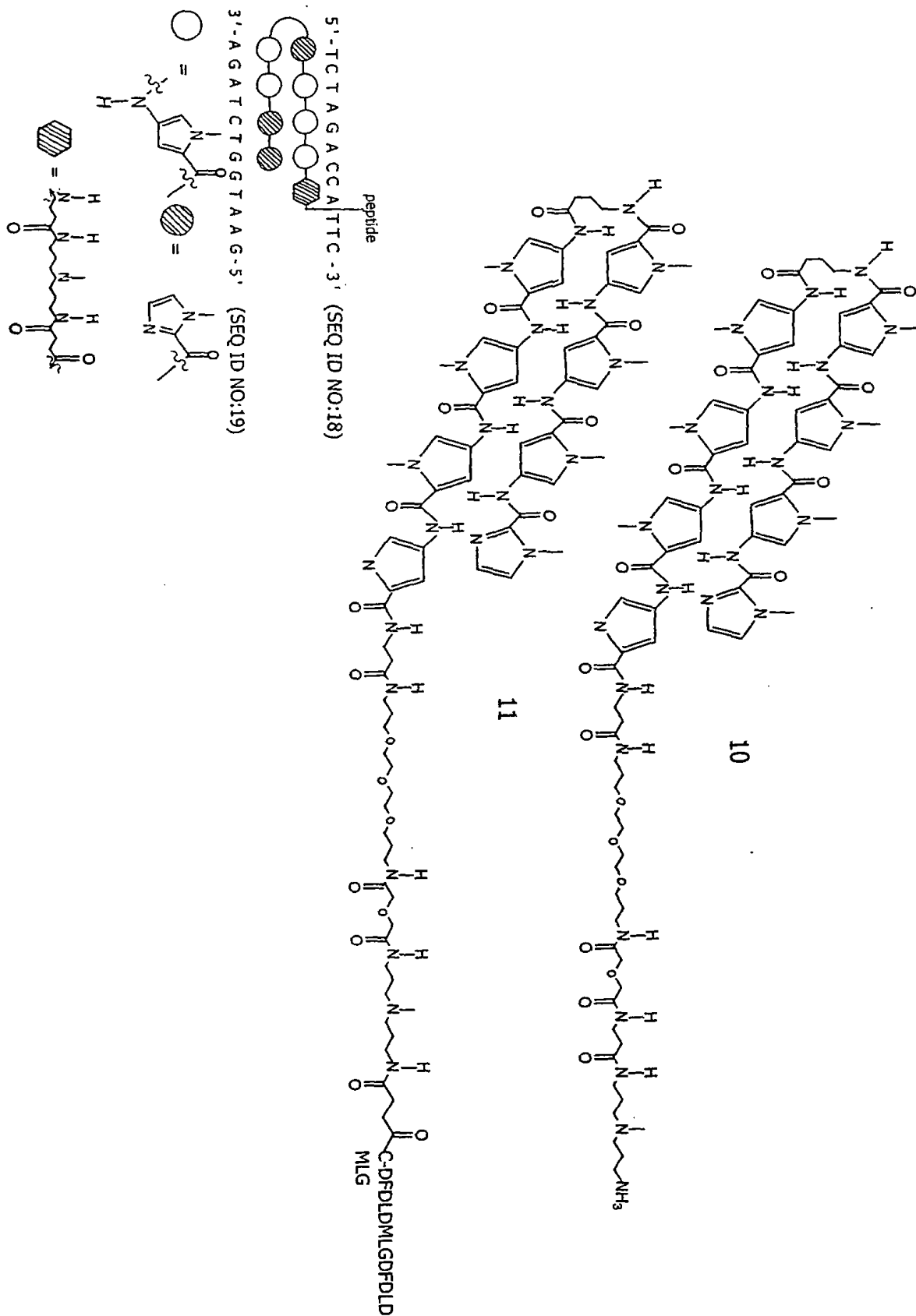
txn 42 quant



1 2 3 4 5 6 7 8 9

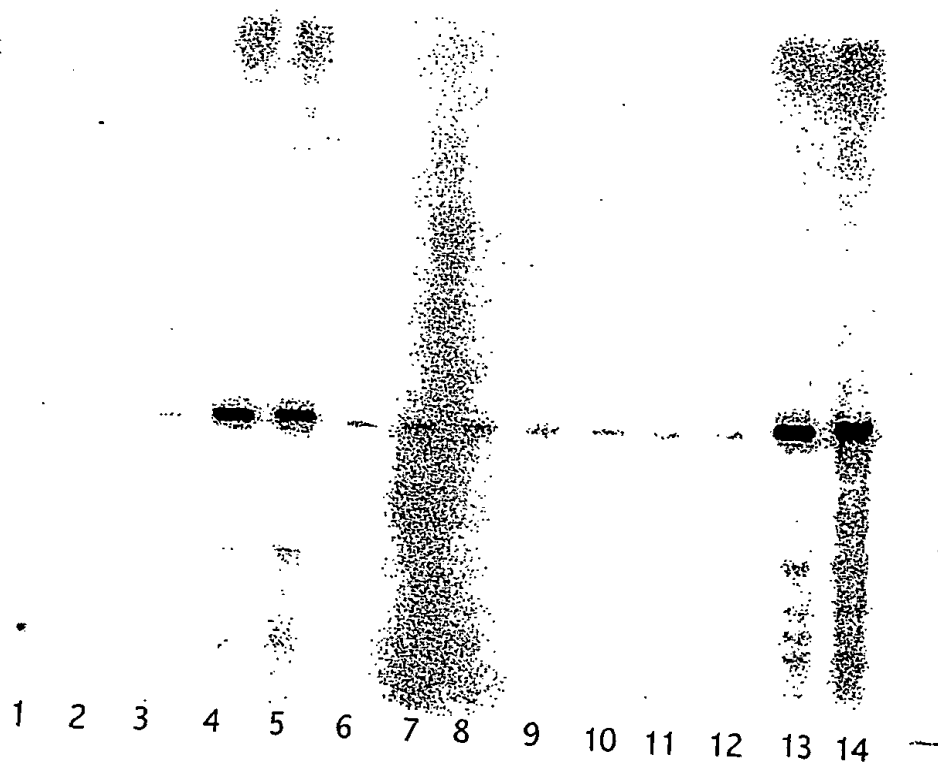
SUBSTITUTE SHEET (RULE 26)

16/17



17/17

txn 43 quant



SUBSTITUTE SHEET (RULE 26)

INTERNATIONAL SEARCH REPORT

In **onal Application No**
PCT/US 00/29617

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 A61K47/48 A61K49/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

CHEM ABS Data, EMBASE, BIOSIS, EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	MAPP, ANNA K. ET AL: "Activation of gene expression by small molecule transcription factors" PROC. NATL. ACAD. SCI. U. S. A. (2000), 97(8), 3930-3935, - 11 April 2000 (2000-04-11) XP002182589 abstract; figures 1,4	1-51
X,P	MAPP, A. K. ET AL: "Preparation of thioesters for the ligation of peptides with non-native substrates" TETRAHEDRON LETT. (2000), 41(49), 9451-9454, XP002182590 See scheme 1,2 abstract	1-29

-/--

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *G* document member of the same patent family

Date of the actual completion of the international search

13 November 2001

Date of mailing of the international search report

19/12/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Berte, M

INTERNATIONAL SEARCH REPORT

In International Application No

PCT/US 00/29617

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	OAKLEY, MARTHA G. ET AL: "Evidence that a minor groove-binding peptide and a major groove-binding protein can simultaneously occupy a common site on DNA" BIOCHEMISTRY (1992), 31(45), 10969-75 , XP002182591 abstract ---	1-29
P, X	LU, XIANGYANG ET AL: "An artificial transcriptional activating region with unusual properties" PROC. NATL. ACAD. SCI. U. S. A. (2000), 97(5), 1988-1992 , XP002182592 abstract ---	1-29
Y	GOTTESFELD, JOEL M. ET AL: "Regulation of gene expression by small molecules" NATURE (LONDON) (1997), 387(6629), 202-205 , XP002182603 page 1 -----	1-29

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims 39,41 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Although claims 42,44-45 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.